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

JIS

第3卷
2024
第 6 期

智能社会研究

第 3 卷
2024
第 6 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 209246 11>

定价：45.00 元

智能社会研究

（双月刊）

2024 年
第 6 期
2024 年 11 月 10 日出版

2022 年 11 月 10 日创刊

总第 13 期

目　次

论文

数字素养对低龄老年人幸福感的影响研究

——基于 CFPS 2020 年数据的实证分析 …… 宋思瑶　贾开阳（ 1 ）

通用人工智能著作权问题的法律因应 …… 罗　艺　曹子贤（ 26 ）

人工智能生成技术方案可专利性界定路径 …… 宋红松　王瑞新（ 48 ）

研究报告

上海数字适老 2.0 探索

——构建新旧媒介并行的微粒社会 …… 张　林（ 64 ）

译文

人工智能中的人的形象

——与约瑟夫·魏岑鲍姆对话

…… 伯恩哈德·珀克森　著　王立秋　译（ 88 ）

计算的限度

——约瑟夫·魏岑鲍姆与聊天机器人 ELIZA

…… 大卫·贝里　著　王立秋　译(107)

从约瑟夫·魏岑鲍姆到 ChatGPT

——与令人倾倒的 AI 技术的批判交锋

…… 克里斯蒂安娜·弗洛伊德　著　王立秋　译(137)

书评

数字平台帝国化的制度演进

——评莱顿维塔的《云端帝国:数字平台如何重塑经济与世界》

…… 牛一帆(169)

技术对社会的再结构

——从卡斯特的网络社会到智能社会 …… 叶莹菲(188)

CONTENTS

THESES

A Study of the Impact of Digital Literacy on the Well-Being of the Under-Aged Older Adults: Empirical Analysis Based on CFPS 2020 Data ··· Song Siyao, Jia Kaiyang(1)

Legal Cause of the Copyright Issue of General Artificial Intelligence ····· Luo Yi, Cao Zixian(26)

Pathway to Defining Patentability for AI-Assisted Inventions ····· Song Hongsong, Wang Ruixin(48)

RESEARCH REPORT

Exploring Digital Aging 2. 0 in Shanghai: Constructing a Granular Society with Parallel New and Old Media ····· Zhang Lin(64)

TRANSLATED TEXTS

The Image of Man in Artificial Intelligence: A Conversation with Joseph Weizenbaum ····· written by B. Pörksen; trans. by Wang Liqiu(88)

The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot ····· written by D. Berry; trans. by Wang Liqiu(107)

From Joseph Weizenbaum to ChatGPT ····· written by C. Floyd; trans. by Wang Liqiu(137)

BOOK REVIEWS

The Institutional Evolution of the Digital Platform Empire: A Review of Vili Lehdonvirta's *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control* ..... Niu Yifan(169)

The Restructuring of Society by Technology: From Manuel Castells' Network Society to the Intelligent Society ..... Ye Yingfei(188)

数字素养对低龄老年人幸福感的影响研究

——基于 CFPS 2020 年数据的实证分析

宋思瑶　贾开阳*

摘要:在中国快速老龄化与数字技术普及的背景下,低龄老年人面临着数字鸿沟的挑战,其数字素养对幸福感的影响愈发显著。本文借助 2020 年中国家庭追踪调查(CFPS)数据,深入探究了低龄老年人数字素养与幸福感之间的关系。研究结果显示,数字素养对低龄老年人的幸福感有着显著的正向影响,尤其是数字社交素养和数字工作素养的影响更为突出。进一步的分析发现,子女数量较多以及处于高网民规模地区的老年人群体,其数字素养对幸福感的促进作用更为明显。机制分析表明,数字素养通过提升社会经济地位和改善主观人际关系判断来增强老年人的幸福感。因此,政策制定者在推动老年人数字素养提升时,应充分考虑地区差异,并加大对低网民规模地区老年人群体的数字教育支持,助力老年人在社交和工作场景中更好地运用数字技术,进而提升幸福感。

关键词:低龄老年人　数字素养　幸福感

一、引言

我国正在经历快速老龄化进程。有学者预测,到 2050 年,中国 65 岁及以上的老年人口将从 2020 年的 1.72 亿人增至 3.66 亿人,占总人口的比例将从 12.0%上升至 26.0%(Feng, Glinskaya & Chen et al., 2020)。尽管我

* 宋思瑶,哈尔滨工程大学人文社会科学学院。贾开阳,哈尔滨工程大学人文社会科学学院。

国老年人使用互联网的比例逐年提升,从 2020 年的 11. 3%增至 2024 年的 14. 3%(CNNIC,2022;CNNIC,2024),但增长速度依然较慢。第 54 次《中国互联网络发展状况统计报告》指出,60 岁及以上的老年人群体是非网民的主要构成部分;截至 2024 年 6 月,该年龄区间的非网民比例高达 62. 0%。这表明,我国老年人在互联网普及和数字技能掌握方面仍面临较大挑战,仍处于信息时代边缘,难以充分享受数字化带来的便利和服务。

2022 年,中央网络安全和信息化委员会印发的《提升全民数字素养与技能行动纲要》明确指出,要推动数字教育资源、数字技能培训、数字产品和信息服务等高质量发展和开放共享,以不断提高人民群众的获得感、幸福感、安全感。2024 年,中央网信办、民政部、全国总工会等 14 个部门联合印发的《2024 年全民数字素养与技能提升月工作方案》提出,要聚焦老年人面临的数字鸿沟问题,助力老年人更好融入数字生活。由此可见,在信息化时代,个人的数字素养与幸福感有着密不可分的联系,个人提升数字素养能更好地适应社会发展和社会生活,从而提升自身幸福感。

老年人群体所面临的数字鸿沟挑战,不仅限制了他们享受数字时代带来的便利,也对其幸福感产生了负面影响,这对于积极应对人口老龄化及数字化社会的高质量发展造成了显著影响。在广泛的老年人群体中,低龄老年人作为数字化应用的主力军(李鸿渊、孙莉莉,2024),较高的数字素养使其能够更好地适应数字化转型,更充分利用数字技术带来的便利。正因如此,低龄老年人群体的数字素养在推动其社会参与、提升生活质量以及增强幸福感方面起着至关重要的作用。据此,本文将集中探讨低龄老年人的数字素养与幸福感之间的关系,旨在揭示低龄老年人数字素养与其自身幸福感之间的影响机制。

二、文献综述

（一）老年幸福感相关的研究

幸福感的研究历史悠久。早在19世纪初，英国经济学家边沁就将幸福量化为效用，以快乐与痛苦的增减为衡量标准，提出“最大多数人的最大幸福”是判断行为和决策的核心依据。20世纪50年代，欧美经济学、心理学学者开始对幸福感这一议题展开深入研究和讨论。随着研究的推进，社会学领域的学者也从多个维度剖析幸福感的影响因素，这些因素可以分为主观和客观两个方面。

从主观层面来看，幸福感通常与个人的心理特质、心理调节能力以及应对生活的方式密切相关。有研究表明，性格乐观的老年人更善于积极应对晚年生活中的各种挑战，如丧偶和孤独，从而维持较高的幸福水平（Diener，Suh，& Lucas et al.，1999）。自尊作为另一个关键心理特质，在老年人晚年生活适应中发挥着重要作用，自尊水平高的老年人更容易获得心理上的满足感和幸福感（章燕敏，2011）。此外，较好的自我效能感也能对老年人自身的幸福感产生积极影响（高锋剑、陈有国、刘培朵等，2017）。积极的心理特质与自我认同感能够帮助老年人应对生活中的挑战，而负面的情绪体验则会削弱老年人的幸福感。有学者（高雨薇、符永川、韩雅馨等，2021）研究表明，孤独感会对老年人的幸福感产生显著的负面影响，要想提高老年人的幸福感，就要降低其孤独感，同时为其提供多方面的社会支持。

老年人幸福感的获得还受到多种客观因素的影响，这些因素大致可以归纳为个人因素和社会环境因素。从个人因素来看，健康状况是决定幸福感的核心变量之一，身体健康、无疾病的老年人通常感到更加幸福（彭馨

瑶、孙睿,2021)。此外,教育背景和经济状况也在幸福感中扮演着重要角色。高学历的老年人由于拥有更强的认知能力和适应性,往往在生活中感受到更多的满足感(王福兴、徐菲菲、李卉,2011)。同时,较高的经济收入为老年人提供了更好的生活保障和安全感,从而显著提升了其幸福感水平(王晓慧,2021)。从社会环境因素来看,家庭支持是老年人情感支持的主要来源,与子女或配偶共同居住的老年人往往因为情感上的联系和陪伴而感到更幸福(张玉银、郑军、范平,2007;沈可、程令国、魏星,2013)。有配偶的老年人相较于丧偶者幸福感更高,子女的照料和陪伴能有效缓解老年人的孤独感,增强其幸福体验(程利娜,2013;高歌、高启杰,2011)。此外,社会支持的数量和质量对幸福感也有显著影响。无论是客观支持还是主观感受到的支持,都能帮助老年人减轻生活压力,增加其对生活的满意度和幸福感(王枫、王茜、庄红平等,2010;任杰、金志成、杨秋娟,2010)。

(二) 数字素养的相关研究

数字素养这一概念最早由吉尔斯特(Gilster, 1997)提出,他认为数字素养意味着个体能够识别和评估数字信息的价值,并能在信息技术环境中有效工作。随着互联网的快速发展和信息社会的形成,数字素养逐渐从单纯的技术技能发展为更复杂的信息处理能力。利文斯顿(Livingstone, 2009)进一步丰富了这一概念,指出数字素养应涵盖如何评估和理解信息,特别是对信息来源的判断和对信息内容的批判性分析。霍布斯(Hobbs, 2010)强调,数字素养必须包括对信息检索、评估、使用和创作的能力,这使得数字素养成为能够批判性思考和理解信息内容的能力。数字素养的概念逐渐丰富,不再仅限于检索和使用信息的能力,还包括内容创作、伦理意识、信息管理等多方面的能力。对此,麦克唐纳等人(MacDonald & Friesen, 2011)提出,数字素养不仅指对技术的熟练掌握,更是一种跨学科的能力,涉

及技术操作、批判性评估、社会文化理解等多个维度。阿拉-穆特卡(Ala-Mutka, 2011)进一步提出,数字素养可以从多个维度来考察,并提出了七个核心维度:访问、评估、管理、分析、创造、交流和批判性思维。

随着相关研究的不断深入,人们对数字素养内涵的认识愈发清晰和全面,越来越多的学者开始探讨数字素养对人们日常生活的影响。有学者发现,数字素养对农村低收入群体的幸福感影响更为显著(Wang, Liu & Cai, 2022),而孤独感较高的人更不倾向于追求较高的数字素养(Ko, Lee & Kim, 2022)。除了对个体幸福感的影响外,也有学者从经济层面展开讨论,认为数字素养能够显著影响个人的财务健康(Choung, Chatterjee & Pak, 2023)。对于女性而言,拥有较高数字技能的女性可以更加安全地上网,并有更多机会学习、创新,为经济建设做出贡献(James, 2022)。不仅如此,数字素养还是获得工作机会的重要条件,尤其是对于农村未接受教育的女性而言(陈丽、翁贞林,2024)。对于老年人来说,互联网的使用可能存在"学习效应",即互联网通过提供信息和学习机会而提高防骗能力,因此数字素养对老年人至关重要(雷晓燕、沈艳、杨玲,2022)。较好的数字素养能够帮助老年人更好地识别虚假信息,从而提升老年人的信息辨识能力(岳小双,2023)。

(三) 数字素养与幸福感

数字技术的迅猛发展,特别是互联网经济的蓬勃兴起,正在深刻地改变社会各个领域的运作模式。从信息获取、社交互动到消费习惯的转变,数字技术已经成为推动经济增长和社会变革的关键力量。随着数字技术日益深入日常生活和生产活动,其对个人行为、社会结构及经济体系的影响也愈发深远,促使人们的生活方式、工作模式以及社会交往方式发生了前所未有的变化。根据第 54 次《中国互联网络发展状况统计报告》,截至 2024 年 6 月,

我国网民数量接近 11 亿人,相比 2023 年 12 月增加了 742 万人,互联网普及率达到了 78.0%。这组数据表明,我国的数字化进程已经取得了显著成效。随着数字技术的广泛普及,人们在日常生活中越来越依赖数字平台,因此数字素养作为一种核心能力变得愈发重要。相较于国外,我国在数字素养领域的研究起步较晚,国内首次明确提出“数字素养”这一概念是在 2006 年(陈怡君,2023)。2022 年 6 月,习近平主席在全球发展高层对话会上再次强调了提高数字素养和技能的重要性,倡导通过增强数字时代的互联互通来推动工业化转型,为国家的发展注入新的动力。

随着国家对互联网经济及数字素养日益重视,许多学者也开始研究互联网使用与个人幸福感之间的影响机制。研究表明,互联网使用能有效提升老年人的幸福感(鲁元平、王军鹏,2020),且不同的互联网使用方式对老年人幸福感的影响各异,利用互联网进行社交和娱乐活动对其幸福感的正向影响最为显著(彭希哲、吕明阳、陆蒙华,2019)。还有学者探讨了互联网使用和幸福感之间的影响机制,得出孤独感在互联网使用与幸福感之间具有中介效应的结论(李红玉,2023)。

综合上述文献回顾,可发现学界目前对数字素养与老年人幸福感相关关系的研究虽取得了一定进展,但仍存在以下两个主要问题:其一,大部分研究对数字素养的测量较为宽泛、模糊,不能真正反映数字素养的时代内涵,因而无法准确探讨数字素养与老年人幸福感之间的关系;其二,在探讨数字素养对老年人幸福感的影响时,较少学者关注不同特征老年人群体中这一影响机制的差异。

因此,本文将运用中国家庭追踪调查(China Family Panel Studies, CFPS)数据,从数字学习素养、数字社交素养、数字工作素养、数字生活素养等方面测量数字素养,借助量化研究方法进一步探讨数字素养与低龄老年人幸福感之间的关系及其在不同特征老年人群体内的差异,以丰富相关实证研究。

三、理论分析与研究假设

老年人幸福感提升的关键,在于社会活动参与(陈佩,2022)。通过社会活动参与,老年人可以提高社会地位、财富状况和交际能力,进而提升其幸福感。社会活动参与涵盖线上活动和线下活动。然而,老年人数字素养普遍偏低。由于身体机能衰老退化以及受教育水平有限,相当一部分老年人面临数字融入困难(袁满,2024),大多数老年人缺乏提高数字素养的机会,在数字时代处于劣势地位。数字素养高的低龄老年人能够较为熟练地使用数字设备进行社交、上网、共享资源、交流信息和协作等,从而提升生活便利性、增强信息获取能力、拓展社交活动、丰富娱乐方式,积极应对老龄化带来的挑战。当前,数字时代的快速发展和广泛应用对老年社会生活产生了深刻影响。数字素养作为数字时代的基础要素,反映了个体对数字时代的适应深度,是老年人享受数字红利、实现积极老龄化的重要前提。老年人可以通过使用手机、电脑等数字设备,依托互联网社交媒体培养自身获取信息资源的能力,增强社会参与感,改善人际互动,提高生活满意度,提升健康管理能力,增强信息甄别能力,进而实现积极老龄化,提升幸福感。

在理论层面,本文将丰富和拓展数字素养与心理健康领域的研究成果,为理解数字素养如何影响老年人幸福感提供新的视角和实证支持,进而构建更加完善的数字素养与心理健康之间的理论框架。在实践层面,本文的研究发现可为政策制定者提供有益参考,助其制定针对性的数字素养培训计划和干预措施,提高老年人的生活质量和社会参与度。此外,本文还将关注不同社会经济背景、教育水平及地区差异下,数字素养对低龄老年人幸福感的影响是否存在显著差异,这将有助于我们更全面地理解数字素养在老年人幸福感中的作用,并为制定更加精准的政策措施提供实证依据。本文

旨在通过 CFPS 数据,实证分析数字素养对低龄老年人幸福感的影响。CFPS 数据提供了丰富的家庭和个人层面信息,为探讨这一问题提供了可靠的数据支持。通过构建合适的计量经济模型,本文将分析数字素养如何影响低龄老年人的幸福感,并探讨其中的作用机制。探讨提升老年人数字素养的路径,对于增进老年人福祉、构建和谐社会具有重要意义。据此,本文提出以下假设:

H1:总体上,低龄老年人的数字素养对其幸福感产生正向影响。

随着信息技术的迅猛发展,数字素养已成为影响老年人生活质量的重要因素。越来越多的研究表明,低龄老年人的数字素养与其幸福感之间存在显著的正向关系。蒋珊珊、宗占红(2024)在一项实证研究中指出,数字融入不仅直接提升了老年人的幸福感($b=0.120$,$p<0.001$),而且通过促进社会参与间接提高了老年人的幸福感。具体而言,数字融入使老年人更容易参与到社会活动中,这种参与感显著提升了他们的生活满意度和幸福感。因此,提升老年人的数字素养,促进他们的数字融入,能够有效提高其幸福感。罗强强等人(罗强强、郑莉娟、郭文山等,2023)的研究结果进一步支持了这一观点。该研究发现,数字素养的提高显著增强了老年人的数字获得感,而数字社会参与在此过程中发挥了中介作用。具体而言,老年人具备较高的数字素养,能够更好地利用数字技术获取信息和服务,从而提升其生活质量,增强幸福感。这一发现强调了数字素养在老年人生活中的重要性,表明数字素养的提升不仅是个人能力的体现,更是增强幸福感的重要途径。袁满(2024)从需求视角进一步论证,老年人对数字技术的需求日益增长,数字素养的提升能够帮助他们更好地适应数字社会,享受更高质量的生活。研究提出,通过教育和培训等方式提升老年人的数字素养,不仅能够满足他们的生活需求,还能够增强他们的社会参与感和归属感,从而进一步提升其幸福感。

H2:社会经济地位和人缘判断是数字素养影响幸福感的重要传导路径。

在探讨数字素养对幸福感的影响时,社会经济地位和主观社会阶层的作用不容忽视。在一项实证分析中,刘晓柳、王俊秀(2020)揭示了社会经济地位和主观社会阶层对个体幸福感的显著影响。研究表明,个体的客观经济状况和他们对社会阶层的主观认知均能预测其幸福感水平,且主观社会阶层感知在幸福感评估中扮演着重要角色。黄婷婷等人(黄婷婷、刘莉倩、王大华等,2016)进一步指出,社会地位比较对幸福感的影响在不同年龄群体中存在差异。这表明,个体在社会中的位置感知,即计量地位,同样能对幸福感产生影响,且这种影响因年龄而异。因此,数字素养的提升可能通过改善个体的社会经济地位和增强其社会地位感知,进而正向影响其幸福感。这些研究为理解数字素养与幸福感之间的关系提供了新的视角,并强调了在提升老年人数字素养时,应考虑其社会经济地位和主观人缘判断的影响。

H3:数字素养提升对低龄老年人幸福感的促进作用在子女数量更多的群体中更强。

数字素养的提升对低龄老年人的幸福感具有显著的促进作用,特别是在子女数量较多的群体中,这种影响更为明显。顾磊、徐生菊(2024)指出,子女对老年人使用智能设备的支持,包括经济补助和技术支持,能有效提高老年人的数字素养水平。这种“数字反哺”不仅帮助老年人更好地适应了数字时代,也增强了他们的社会参与感和自我效能感,从而提升了幸福感。袁满(2024)进一步强调,相较于自我学习,老年人更倾向于直接寻求子女的帮助来解决数字技术问题。这种依赖关系在子女数量较多的家庭中更为常见,使得这些老年人在数字技能的获取和应用上更为得心应手,进而在日常生活中体验到更多的便利和乐趣,增强了他们的幸福感。因此,子女的支

持是低龄老年人数字素养提升和幸福感增强的重要传导路径,尤其在子女数量较多的家庭中,这种作用更为显著。

H4a:数字素养提升对低龄老年人幸福感的促进作用在低网民规模地区更明显。

在低网民规模地区,数字素养提升对老年人幸福感的促进作用更明显。杨森雯(2023)指出,网络使用和知识素养对老年人的幸福感暂无显著影响,主要是数字接入发挥着显著作用。这意味着在低网民规模地区,提升老年人的数字接入可能对提升他们的幸福感有更直接的影响。

H4b:数字素养提升对低龄老年人幸福感的促进作用在高网民规模地区更明显。

研究表明,互联网对老年人生活幸福感具有积极影响。老年人接触或使用网络的频率越高,他们对生活的满意程度越高。这表明,在互联网普及率较高的地区,老年人通过互联网提升数字素养,可能更容易感受到数字红利,从而提高他们的幸福感。

综上所述,低龄老年人的数字素养与其幸福感之间存在显著的正向关系。通过提升数字素养,老年人能够更好地融入数字社会,享受数字技术带来的便利和服务,从而提高其生活质量和幸福感。因此,社会各界应重视老年人的数字素养教育,为他们创造更好的数字环境,以促进其幸福感的提升。

四、数据来源、变量说明与模型选择

(一) 数据来源

在本文中,笔者使用了 2020 年度的中国家庭追踪调查数据。该调查项目覆盖了中国 31 个省、自治区、直辖市(不包括香港、澳门、台湾),涵盖了约 15 000 户家庭,具有广泛的调查对象、丰富的调查层次和高质量的数据特点,能够为本文提供具有代表性和可靠性的基础数据。鉴于本文的研究对象聚焦于低龄老年人群体,因此将研究对象的年龄范围限定在 60—70 岁。为确保数据分析的准确性和可靠性,在处理数据时,删除了存在关键变量缺失的样本。最终,符合分析条件的有效样本数量为 998 个。通过这样的筛选过程,确保了数据的有效性和结论的可信度,从而为后续分析提供了坚实的基础。

(二) 变量说明

1. 因变量

本文采用 2020 年 CFPS 问卷中“您觉得自己有多幸福?”这一题项来测度老人的幸福感。该变量的取值范围为 0—10,数值越大,表示幸福感程度越高。

2. 自变量

数字素养。本文参考已有研究(李丽莉、徐嘉、梅燕等,2024)的做法,从数字学习、数字社交、数字工作和数字生活四个维度出发,从 CFPS 数据中整合了 12 项与数字素养相关的指标。为了客观地确定各个指标的权重,本文采用普通熵权法进行计算,最终将各项权重加总,得出数字素养的综合

得分。数字素养的具体指标选取和测量如表 1 所示。

表 1　　数字素养测量

数字素养维度	测量题目	赋值
数字学习素养	是否使用网络学习	是=1;否=0
	网络对学习的重要性	按重要性程度赋值 1—5:非常重要=5;非常不重要=1
数字社交素养	是否使用微信	是=1;否=0
	朋友圈分享频率	从不=1;几个月一次=2;一个月一次=3;一个月两三次=4
	网络对与亲友保持联系的重要性	按重要性程度赋值 1—5:非常重要=5;非常不重要=1
数字工作素养	网络对工作的重要性	按重要性程度赋值 1—5:非常重要=5;非常不重要=1
	互联网对获取信息的重要性	按重要性程度赋值 1—5:非常重要=5;非常不重要=1
数字生活素养	是否玩网络游戏	是=1;否=0
	是否网上购物	是=1;否=0
	是否看短视频	是=1;否=0
	网络对休闲娱乐的重要性	按重要性程度赋值 1—5:非常重要=5;非常不重要=1
	网络对日常生活的重要性	按重要性程度赋值 1—5:非常重要=5;非常不重要=1

3. 控制变量

参考已有研究(李晓静、陈哲、夏显力,2022),本文将控制变量分为个体层面和家庭层面两类:个体层面的控制变量包括性别、受教育程度、婚姻状况、健康状况、户口类型;家庭层面的控制变量包括家庭人口规模、家庭年收入对数以及子女关系。各项变量的描述性统计和说明如表 2 所示。

表 2　　变量说明与描述性统计

	变量名称	定义与赋值	平均值	标准差	最小值	最大值
被解释变量	幸福感	0—10，数值越高，幸福感越强	7.894	1.935	0	10
核心解释变量	数字素养	采用熵权法进行测算	0.520	0.163	0.083	0.958
控制变量	年龄	调查年份的年龄（岁）	64.377	3.030	60	70
	性别	男=1；女=0	0.544	0.498	0	1
	受教育程度	文盲/半文盲=0；小学=1；初中=2；高中/中专/技校/职高=3；大专及以上=4	1.891	1.209	0	4
	婚姻状况	已婚=1；未婚=0	0.908	0.290	0	1
	健康状况	非常健康=1；很健康=2；比较健康=3；一般=4；不健康=5	2.724	1.178	1	5
	户口类型	非农户口=0；农业户口=1	0.388	0.488	0	1
	家庭人口规模	被访者核心家庭成员总数	1.364	4.828	1	15
	家庭收入	过去12个月家庭总收入的对数	10.945	1.114	6.908	13.989
	子女关系	很不亲近=1；不大亲近=2；一般=3；亲近=4；很亲近=5	4.504	0.660	2	5

（三）模型选择

由于幸福感是一个定序变量，本文构建了有序概率模型（ordered probit model）来估计数字素养对居民幸福感的影响。具体的模型设定为：

$$happiness_i = \alpha + \beta \cdot digital_literacy_i + \gamma \cdot X + \varepsilon$$

在该模型中，因变量 $happiness_i$ 是个体报告的幸福感水平；自变量 $digital_literacy_i$ 表示个体 i 的数字素养水平；X 是控制变量集合，包括个体层面

的因素(如年龄、性别、教育水平)以及家庭层面的因素;ε 是随机扰动项,代表无法观测的个体差异和外部因素,假设其服从标准正态分布。

五、实证分析

(一) 基准回归分析

表 3 报告了使用 Probit 模型估计数字素养对低龄老年人幸福感的影响结果。模型(1)是未加入任何控制变量的模型,模型(2)则加入了个体及家庭层面的控制变量。在模型(1)中,我们首先考察了数字素养对幸福感的影响。回归结果显示,数字素养的系数为 0.997,且在统计上显著($p<0.01$),表明数字素养的提高与幸福感之间存在显著的正向关系。在模型(2)中,进一步引入了控制变量,包括年龄、性别、教育水平等。结果表明,数字素养的系数增至 1.125,同样显著($p<0.01$)。这一结果表明,控制了其他背景因素后,数字素养对幸福感的影响依然显著,且其影响力度有所增强。进一步的边际效应分析显示,数字素养每提升一个单位,居民报告更高幸福感的概率将增加 112.5%。

在模型(3)、模型(4)、模型(5)和模型(6)中,分别将数字学习素养、数字社交素养、数字工作素养和数字生活素养作为核心自变量,系统地考察了各个维度的数字素养对低龄老年人幸福感的影响。通过表 3 可以看出,模型(3)、模型(4)、模型(5)和模型(6)均在 1%的统计水平上显著正向影响老年人的幸福感。其中,数字社交素养对幸福感的影响最大,系数为 4.224,表明社交互动和在线沟通能力的提升显著促进了个体的幸福感。数字工作素养紧随其后,系数为 4.208,也对幸福感产生了显著的正向影响。相比之下,数字学习素养的影响较小,系数为 2.972,而数字生活素养的影

响最小,系数为 1.019,但仍显示出显著的正向效应。

从控制变量来看,婚姻状况、健康状况和子女关系是影响幸福感最重要的控制变量,它们在所有模型中均表现出显著的正向效应。婚姻状况和健康状况对幸福感的影响较为强烈,而教育程度、户口类型则显示了负向影响,反映了教育水平和城乡差异对幸福感的潜在影响。其他控制变量,如年龄、性别、家庭收入和家庭人口规模,对幸福感的影响较小或不显著。

表 3　　基准回归结果

解释变量	被解释变量:幸福感					
	模型(1)	模型(2)	模型(3)	模型(4)	模型(5)	模型(6)
数字素养	0.997*** (4.271)	1.125*** (4.429)				
数字学习素养			2.972*** (2.989)			
数字社交素养				4.224*** (3.742)		
数字工作素养					4.208*** (4.821)	
数字生活素养						1.019** (2.287)
年龄		−0.003 (−0.245)	−0.007 (−0.515)	−0.009 (−0.614)	−0.008 (−0.586)	−0.003 (−0.195)
性别		0.054 (0.640)	0.060 (0.714)	0.045 (0.536)	0.038 (0.454)	0.071 (0.844)
受教育程度		−0.106** (−2.449)	−0.107** (−2.477)	−0.107** (−2.470)	−0.103** (−2.382)	−0.091** (−2.126)
婚姻状况		0.461*** (3.081)	0.416*** (2.79)	0.414*** (2.777)	0.467*** (3.119)	0.437*** (2.923)

（续表）

解释变量	被解释变量：幸福感					
	模型（1）	模型（2）	模型（3）	模型（4）	模型（5）	模型（6）
健康状况		0.171***	0.173***	0.170***	0.168***	0.172***
		（4.694）	（4.741）	（4.654）	（4.591）	（4.726）
户口类型		−0.207**	−0.234**	−0.188*	−0.258**	−0.186*
		（−2.031）	（−2.277）	（−1.844）	（−2.506）	（−1.818）
家庭人口规模		−0.020	−0.028	−0.019	−0.024	−0.026
		（−0.806）	（−1.127）	（−0.751）	（−0.947）	（−1.024）
家庭收入		−0.023	−0.002	−0.016	−0.009	−0.008
		（−0.456）	（−0.043）	（−0.324）	（−0.177）	（−0.167）
子女关系		0.301***	0.303***	0.306***	0.293***	0.306***
		（4.898）	（4.917）	（4.965）	（4.758）	（4.971）
Pseudo R^2		0.036	0.032	0.034	0.038	0.030

注：* $p<10\%$；** $p<5\%$；*** $p<1\%$。括号内为稳健标准误。下同。

（二）内生性处理

在本文中，数字素养与幸福感之间可能存在潜在的逆向因果关系，即幸福感较高的个体可能更倾向于提升自己的数字素养。此外，考虑到研究中可能遗漏了一些关键变量，这些未被纳入模型的变量可能导致估计偏误。例如，社会经济地位、健康状况或社会支持等因素可能同时影响个体的数字素养和幸福感，从而导致遗漏变量偏误。逆向因果和遗漏变量偏误是内生性问题产生的主要原因，这将直接影响回归结果的可靠性。因此，为了有效解决这一内生性问题，本文采用了工具变量法进行进一步分析，通过找到与数字素养相关但与幸福感仅通过数字素养相关的工具变量，消除因果关系中的内生性偏误，并获得更为可靠的估计结果。

本文采用手机短信重要性作为工具变量。选择这一变量的原因，一方

面是手机短信的使用和依赖通常与数字素养之间具有较强的正相关性。随着信息技术的发展,手机短信作为一种便捷的沟通方式,在现代社会中扮演着重要角色。通常,频繁使用手机短信的人往往具备较高的数字素养,因为他们需要掌握手机操作、信息管理等技术技能,这些技能的提升与数字素养水平紧密相关。另一方面,手机短信的重要性对老年人幸福感的影响主要是通过数字素养间接发生的。良好的手机短信使用能力不仅能帮助个体保持社交联系,还能促进老年人的信息获取并提高生活便捷度,从而间接提升老年人的幸福感,这一机制符合外生性条件。我们首先对工具变量的可靠性进行了检验,Cragg-Donald Wald F 统计量均大于 10,可以认为本文选取的变量不存在弱工具变量问题。

表 4 汇报了使用工具变量法的回归结果。在第一阶段,手机短信重要性与数字素养之间的关系显著,系数为 0.115(p<0.05)。进入第二阶段后,数字素养对幸福感的影响显著,系数为 2.675(p<0.01),这表明在考虑了模型可能存在的内生性问题之后,数字素养仍然对老年人的幸福感具有显著的正向影响。

表 4　　工具变量法的回归结果

解释变量	2SLS	
	第一阶段	第二阶段
数字素养		2.675*** (2.274)
手机短信重要性	0.115** (2.239)	
控制变量	已控制	已控制
一阶段 F 统计值	100.473	
Wald 检验		0.021

（三）稳健性检验

幸福感通常被视为个体对当前生活的主观评估，它在一定程度上反映了个人的生活满意度。因此，已有研究将生活满意度作为衡量幸福感的替代指标（Helliwell，Huang & Grover et al.，2018）。为了进一步检验研究结果的稳健性，本文将因变量替换为生活满意度，并重新评估数字素养的影响。生活满意度这一变量来源于 CFPS 2020 中的一个问题："您对自己生活的满意度？"该问题采用 5 点量表，1 表示"很不满意"，5 表示"非常满意"。

由于数字素养在不同人群和年龄段之间存在显著差异，尤其是在性别和年龄方面，个体的数字技能往往受社会角色、教育背景和技术接触机会等因素的影响。因此，为了确保本研究结果的稳健性，并减少潜在的异质性影响，本文决定剔除部分特定样本进行稳健性检验。具体而言，鉴于男性和女性在数字技术的使用模式与技术接触方式方面可能存在差异，且老年人群体（尤其是 65 岁以上的个体）在数字素养上的表现可能与其他年龄段显著不同，因此本研究将剔除女性样本以及年龄大于 65 岁的样本，重新估计模型。

表 5 报告了稳健性检验的结果。模型（7）是替换了因变量生活满意度的回归模型，而模型（8）、模型（9）则是剔除了女性样本以及年龄大于 65 岁的样本的回归模型。可以看到，这些估计结果与基准回归结果基本一致，表明前文实证结果具有稳健性。

表 5　　稳健性检验结果

解释变量	模型（7）	模型（8）	模型（9）
数字素养	0.485** （1.799）	1.149*** （3.389）	1.214** （2.509）
控制变量	已控制	已控制	已控制
Pseudo R^2	0.062	0.043	0.067

（四）影响机制分析

虽然已有研究证实了数字素养对老年人幸福感具有显著的正向影响，但其内在作用机制仍需进一步探讨。本文将通过社会经济地位和主观人缘判断这两个维度，深入分析数字素养对老年人幸福感提升的机制。

表 6 报告了数字素养对社会经济地位和主观人缘判断的影响结果。结果显示，数字素养对两者均具有显著的正向影响。具体而言，数字素养每提高一个单位，社会经济地位提高 0.122 个单位，且在 1%水平上显著；数字素养每提高一个单位，主观人缘判断提高 1.120 个单位，且在 1%水平上显著。这表明，数字素养的提升有助于老年人获得更高的社会经济地位和更好的社会关系，从而为其幸福感的提升奠定基础。

表 6　　数字素养对社会经济地位及主观人缘判断的影响

解释变量	因变量：社会经济地位	因变量：主观人缘判断
数字素养	0.122*** （3.186）	1.120*** （5.488）
控制变量	已控制	已控制
Pseudo R^2	0.0056	0.0084

为了更好地验证社会经济地位及主观人缘判断的中介效应，本研究采用 KHB 方法进行测算。表 7 的结果显示，数字素养对幸福感的总效应系数为 1.976，表明数字素养显著正向影响老年人的幸福感。此外，间接效应系数为 1.167，即数字素养通过提高社会经济地位和改善主观人缘判断对幸福感的间接影响，中介效应的贡献率达到 59.05%。这也凸显了社会经济地位和主观人缘判断在数字素养与老年人幸福感关系中的重要作用。

表 7 中介检验结果(KHB)

项目	系数	标准误	Z
总效应	1.976***	0.373	5.290
直接效应	0.810***	0.390	2.070
间接效应	1.167***	0.258	4.520
中介效应贡献率	59.05%		

(五)异质性分析

为了深入探讨数字素养对老年人幸福感的影响,本文将实施一项异质性分析。具体而言,本文将老年人群根据其子女数量进行分类,分为无子女、仅有一个子女、两个子女以及两个以上子女的不同子群体。此外,本文依据第 54 次《中国互联网络发展状况统计报告》所提供的各省网民规模数据,将老年人所在地区划分为低网民规模地区与高网民规模地区。表 7 中的模型(10)和模型(11)分别为低网民规模地区与高网民规模地区的分组回归结果。结果显示,在控制了相关影响因素后,在高网民规模地区,数字素养对老年人幸福感的促进作用更为明显。模型(12)、模型(13)、模型(14)、模型(15)分别为不同子女数的回归结果。从回归结果可以看出,在没有子女的老年人群体中,回归结果并不显著,这表明对于没有子女的老年人,数字素养与幸福感之间的关系不明确。而对于有子女的老年人来说,随着子女数量的增加,数字素养对幸福感的正向影响趋于显著。

表 8 异质性分析结果

解释变量	模型(10)	模型(11)	模型(12)	模型(13)	模型(14)	模型(15)
数字素养	0.948** (2.581)	1.322*** (3.706)	1.338 (1.184)	0.857** (2.282)	1.464*** (3.185)	1.628*** (2.513)

（续表）

解释变量	模型(10)	模型(11)	模型(12)	模型(13)	模型(14)	模型(15)
控制变量	已控制	已控制	已控制	已控制	已控制	已控制
Pseudo R^2	0.046	0.034	0.178	0.036	0.054	0.075

六、结论与政策启示

本文通过有序 Probit 模型探讨了数字素养对低龄老年人幸福感的影响，得出了若干重要结论。首先，数字素养与幸福感之间存在显著的正向关系。在基础回归模型中，数字素养的提高显著增加了老年人报告较高幸福感的概率，表明数字素养的提升有助于提高老年人的幸福感。具体来说，数字素养每提升一个单位，老年人报告更高幸福感的概率将增加约 112.5%。进一步分析发现，数字素养的各个维度对幸福感的影响存在差异，其中数字社交素养和数字工作素养对幸福感的促进作用最为显著，分别使老年人幸福感得到了较大幅度的提升；而数字学习素养和数字生活素养的影响相对较小，但依然表现为正向效应。

此外，异质性分析揭示了数字素养对不同子女数和地区老年人群体幸福感的影响差异。具体而言，在没有子女的老年人群体中，数字素养与幸福感之间的关系并不显著，说明子女的陪伴和支持可能是影响这一群体幸福感的重要因素。相比之下，在有子女的老年人群体中，数字素养对幸福感的正向影响愈加显著，且随着子女数量的增加，数字素养对幸福感的提升作用更为明显。这表明，子女的存在和数量在老年人幸福感的提升过程中起到了重要的调节作用。

同时，地区层面的差异也对数字素养与幸福感的关系产生了显著影响。在高网民规模地区，老年人受益于较高的数字素养，其幸福感的提高幅度明

显高于低网民规模地区。这样的差异可能与地区信息化程度、网络使用习惯以及老年人对数字技术的接触机会等因素密切相关。因此,在推动老年人数字素养的提升时,相关政策应考虑地区差异,并加大对低网民地区老年人群体的数字教育支持,以确保老年人在各类社交和工作场景中能够充分利用数字技术,进而提升他们的幸福感。

总体而言,本文表明,数字素养的提升,尤其是在社交和工作领域的提升,对老年人幸福感提升具有重要促进作用,而这一作用受到子女关系和地区差异的调节。

参考文献

CNNIC,2022,《中国互联网络发展状况统计报告(第 50 次)》,https://www. cnnic. net. cn/n4/2022/0914/c88-10226. html。

CNNIC,2024,《中国互联网络发展状况统计报告(第 54 次)》,https://www. cnnic. net. cn/n4/2024/0829/c88-11065. html。

陈丽、翁贞林,2024,《数字素养对农村女性劳动力就业质量影响》,《江西财经大学学报》第 4 期。

陈佩,2022,《老龄化背景下老年人幸福感及影响因素研究——基于 CGSS 2017 数据的实证分析》,《应用数学进展》第 5 期。

陈怡君,2023,《国内数字素养研究综述——基于 CiteSpace 的文献计量与可视化分析》,《昆明理工大学学报(社会科学版)》第 6 期。

程利娜,2013,《社会支持、自我效能感对丧偶老年人主观幸福感的影响》,《公共卫生与预防医学》第 1 期。

高锋剑、陈有国、刘培朵等,2017,《心理韧性、孤独感、自我效能感预测老年人主观幸福感的优势分析》,《心理与行为研究》第 2 期。

高歌、高启杰,2011,《农村老年人生活满意度及其影响因素分析——基于河南省叶县的调研数据》,《中国农村观察》第 3 期。

高雨薇、符永川、韩雅馨等,2021,《延安市老年人孤独感、社会支持与主观幸福感的关系研究》,《延安大学学报(自然科学版)》第 3 期。

顾磊、徐生菊,2024,《老年群体数字素养调查与分析——以江苏省南通市为例》,《文献与数据学报》第 1 期。

黄婷婷、刘莉倩、王大华等,2016,《经济地位和计量地位:社会地位比较对主观幸福感的影响及其年龄差异》,《心理学报》第 9 期。

蒋珊珊、宗占红,2024,《数字融入对老年人主观幸福感影响及路径研究》,《南京医科大学学报(社会科学版)》第 2 期。

雷晓燕、沈艳、杨玲,2022,《数字时代中国老年人被诈骗研究——互联网与数字普惠金融的作用》,《金融研究》第 8 期。

李红玉,2023,《老年人互联网使用特点及其与孤独感和主观幸福感关系研究》,《中国新通信》第 20 期。

李鸿渊、孙莉莉,2024,《数字化应用与低龄老年人生活满意度——基于社区参与的中介效应研究》,《上海对外经贸大学学报》第 5 期。

李丽莉、徐嘉、梅燕等,2024,《数字素养对农民创业决策的影响:来自 CFPS 的经验证据》,《农林经济管理学报》第 1 期。

李晓静、陈哲、夏显力,2022,《数字素养对农户创业行为的影响——基于空间杜宾模型的分析》,《中南财经政法大学学报》第 1 期。

刘晓柳、王俊秀,2020,《社会经济地位与主观社会阶层对幸福感的影响——基于 CGSS 2010—2015 的实证分析》,《广西师范大学学报(哲学社会科学版)》第 5 期。

鲁元平、王军鹏,2020,《数字鸿沟还是信息福利——互联网使用对居民主观福利的影响》,《经济学动态》第 2 期。

罗强强、郑莉娟、郭文山等,2023,《"银发族"的数字化生存:数字素养对老年人数字获得感的影响机制》,《图书馆论坛》第 5 期。

彭希哲、吕明阳、陆蒙华,2019,《使用互联网会让老年人感到更幸福吗?——来自 CGSS 数据的实证研究》,《南京社会科学》第 10 期。

彭馨瑶、孙睿,2021,《机构养老老年人的主观幸福感研究综述》,《财富时代》第 9 期。

任杰、金志成、杨秋娟,2010,《老年人主观幸福感影响因素的元分析》,《中国临床心理学杂志》第 1 期。

沈可、程令国、魏星,2013,《居住模式如何影响老年人的幸福感?》,《世界经济文汇》第 6 期。

王枫、王茜、庄红平等,2010,《老年人主观幸福感及其影响因素分析》,《医学与社会》第 12 期。

王福兴、徐菲菲、李卉,2011,《老年人主观幸福感和孤独感现状》,《中国老年学杂志》第 13 期。

王晓慧,2021,《社会经济地位对老年人主观幸福感的影响研究》,《大连理工大学学报(社会科学版)》第 3 期。

杨森雯,2023,《数字鸿沟与老人主观幸福感》,上海财经大学硕士学位论文。

袁满,2024,《需求视角下城市老年人数字素养提升对策研究》,《可持续发展》第 2 期。

岳小双,2023,《老年人谣言辨识的数字素养提升研究》,四川大学硕士学位论文。

张玉银、郑军、范平,2007,《居住方法对老年人生活质量的影响》,《中国疗养医学》第 4 期。

章燕敏,2011,《国内老年人主观幸福感研究综述》,《大众科技》第 3 期。

Ala-Mutka, K. 2011, *Mapping Digital Competence: Towards a Conceptual Understanding*, DOI:10. 13140/RG. 2. 2. 18046. 00322.

Choung, Y., S. Chatterjee & T. Pak 2023, "Digital Financial Literacy and Financial Well-Being." *Finance Research Letters* 58.

Diener, E., E. Suh & R. Lucas et al. 1999, "Subjective Well-Being: Three Decades of Progress." *Psychological Bulletin* 125(2).

Feng, Z., E. Glinskaya & H. Chen et al. 2020, "Long-Term Care System for Older Adults in China: Policy Landscape, Challenges, and Future Prospects." *Lancet* 396(10259).

Gilster, P. 1997, *Digital Literacy*, New York:Wiley.

Helliwell, J., H. Huang & S. Grover et al. 2018, "Empirical Linkages Between Good Governance and National Well-Being." *Journal of Comparative Economics* 46(4).

Hobbs, R. 2010, *Digital and Media Literacy: A Plan of Action*, Washington, DC.: The Aspen Institute.

James, A. 2022, "Women in the Gig Economy: Feminising 'Digital Labour'." *Work in the Global Economy* 2(1).

Ko, D., J. Lee & H. Kim 2022, "Loneliness, Implicit-Self and Digital Literacy." *Frontiers in Psychology* 13.

Livingstone, S. 2009, "Digital Literacy: The Role of the Media in Digital Literacy." in C. Lankshear & M. Knobel eds., *Handbook of Research on New Literacies*, New York: Routledge.

MacDonald, S. & B. Friesen 2011, *Digital Literacy: A Critical Thinking Approach*, New York: Routledge.

Wang, J., C. Liu & Z. Cai 2022, "Digital Literacy and Subjective Happiness of Low-Income Groups: Evidence from Rural China." *Frontiers in Psychology* 13.

通用人工智能著作权问题的法律因应*

罗　艺　曹子贤**

摘要:通用人工智能标志着强人工智能时代的到来,以 ChatGPT 4.0 为代表的生成式人工智能在应用过程中,其生成物对传统的知识产权规范提出了全新的挑战。本文在阐述通用人工智能及其应用现状的基础上,分析论证了通用人工智能存在的输入端、输出端和生成物归属问题的相关著作权困境,从而构建出通用人工智能著作权问题法律回应的具体路径:应当承认通用人工智能输入端的必要性;发挥版权集体管理和代理制度功能,设立针对通用人工智能大模型训练的作品"禁用期";区分不同情况,考虑赋予人工智能独立或部分版权;将通用人工智能虚拟作者的作品收入进行相应合理分配。通过具体的制度构建,在保障自然人权利的同时,激发通用人工智能的创造性,共同为人类福祉贡献力量。

关键词:通用人工智能　人工智能生成物(AIGC)　回应型法　形式人

人工智能分为专用人工智能(SAI)和通用人工智能(AGI)。专用人工智能是针对特定领域和专业任务而发挥其功能的智能系统,由于其功能单一、模式固定,因此被称为"弱人工智能";通用人工智能的功能范围更加全面,可以适应多种情况与需求,具有自主分析和解决问题的能力,在一定程

* 本文系甘肃省法学会 2024 年度法治甘肃省级课题"新时代弘扬发展'枫桥经验'和'马锡五审判方式'研究"(项目批准号:035)、四川省教育厅人文社会科学重点研究基地四川基层社会风险防控治理研究中心乐山师范学院 2024 年项目资助重点项目"基层社会风险防控治理智能化研究"(项目批准号:JCFXFK24-032A)、甘肃省高级人民法院民族法制文化研究所 2024 年度研究课题"新时代'枫桥经验'视域下源头治理协同化路径研究"[项目批准号:GSFYSKT(2024)25]的阶段性研究成果。

** 罗艺,东华理工大学文法与艺术学院、华东政法大学博士后流动站。曹子贤,甘肃政法大学法学院。

度上超越了人类智力,因此被称为“强人工智能”或“完全人工智能”。党的二十大报告指出,要“加强知识产权法治保障,形成支持全面创新的基础制度”。习近平总书记指出:“要加强人工智能发展的潜在风险研判和防范,维护人民利益和国家安全,确保人工智能安全、可靠、可控。”随着 ChatGPT 4.0、混元助手以及豆包等生成式人工智能的出现,通用人工智能技术得到了迅速发展,从而对我国现有的知识产权保护体系提出了前所未有的挑战,这都需要对法律制度的探索与回应。

一、通用人工智能及其应用现状

(一) 通用人工智能的基本认知

人工智能(artificial intelligence,AI)这个概念诞生于 20 世纪四五十年代。受计算能力不足等多种因素影响,人工智能的发展较为缓慢,仅停留在专用人工智能阶段。到了 21 世纪,人工智能技术得到了跨越式发展,开始适用于自动驾驶、医疗设备以及金融分析等多个领域。2023 年初,以 ChatGPT 为代表的生成式人工智能的出现,标志着人类从专用人工智能时代正式迈入通用人工智能时代(《学术前沿》编者,2023)。通用人工智能相较于专用人工智能,在多方面得到了提升和发展:拥有大规模数据处理和存储能力,通过大量的数据投喂来训练以及优化模型;拥有强大的计算能力以及先进的算法模型,在生成决策结果具备科学性的同时,也保证了效率;拥有自主学习和适应能力,不断适应新的情况和要求,从而不断优化结果;拥有泛化能力,在不同的任务和环境中表现出了具备人类甚至优于人类的智能水平。通过收集、训练、上下文理解、生成、调优等五个环节(张金平,2023),通用人工智能可生成良好的图片、音频、文本等内容:先收集大量

的样本数据进行学习以及预处理,随后采用无监督方式对大模型进行不断的训练;大模型通过用户输入记录进行上下文理解,推断用户的意思需求,从而预生成内容;最后,大模型会对即将生成的内容进行检查优化并生成内容。

(二) 通用人工智能目前的应用模式

目前,通用人工智能的典型表现形式是生成式人工智能。生成式人工智能不仅具有文字生成功能,还具有强大的音频和图像生成功能。生成式对抗网络(GAN)和变分自编码器(VAE)便是具有音频和图片生成功能的生成式人工智能。GAN 由两个神经网络组成,一个用于生成数据,另一个用于判断所生成的数据是否真实。通过这种交互,GAN 能够生成非常真实的数据。VAE 也是一种生成式网络,它通过学习数据中的模式来生成新的、相似的数据。与 GAN 不同的是,VAE 可以生成连续的数据。从纯技术角度来看,生成式人工智能主要由用户生成内容(UGC)和专业生成内容(PGC)两部分构成,其中最能反映生成式人工智能落地效应的是对人类经济、文化产生直接影响的 UGC 领域(何大安、许一帆,2023)。无论是在数字艺术、音乐、电影或其他艺术领域,AI 都已经开始被广泛接纳,而不再仅仅停留在概念阶段。多个商业平台和服务已经开始提供基于这些模型的艺术创作工具,为艺术家提供了更为丰富和多元的创作选择,推动整个艺术产业进入了一个全新的时代(陈抱阳,2023)。

1. 语音模拟应用

现有的 AI 生成模型网站,在使用者通过文字输入、参数调节、素材投喂等操作后,将生成拟真的音频。网站的使用者会通过模型生成模仿明星声音的音频,用以制作短视频,上传至网络平台。所模仿的人声,部分会用于对歌曲的翻唱。例如,Bert-VITS2 AI 便是由谷歌开发的一款基于自然语

言处理(NLP)和计算机视觉(CV)技术的 AI 模型。

2. 绘画生成应用

许多 APP 和网站都具备 AI 绘画生成技术,尽管生成的绘画作品质量参差不齐。生成的基本程序有两种:第一种,使用者选择所需的画风类型,通过文字描述,由 AI 模型提取特征向量,再通过图像编码器生成作品;第二种,使用者向 AI 模型上传现实照片,AI 在提取照片特征后,将照片内容生成为绘画版。例如,用户上传自己的自拍照,AI 会生成一个自己的肖像。如今,许多平台上的生成式人工智能创作者分享着自己通过人机合作所生成的作品。

3. 音乐创作应用

现在的 AI 音乐生成方式分为两类:一类是选择歌曲曲风生成完整的作品,另一类是生成一首完整音乐作品的组成要素。一首完整的音乐作品需要作曲、编曲、混音、母带和人声,而现在的通用人工智能技术可以较好地完成作曲、编曲以及人声的环节。例如,国外的 AI 音乐网站 Soundful,其突出特点是编曲伴奏的生成,用户可以在其中选择不同风格的伴奏,如 POP、EDM 等。

4. 新闻稿、分析报告等文章内容的生成

AI 大模型会将使用者输入的关键信息(如事件、人物、地点等)组合成新闻稿,或者提取一篇有版权的新闻稿的关键信息,通过语言模型套用模板润色,形成一篇新的新闻稿。这种方式在行业领域叫作“洗稿”,而通用人工智能大模型的出现无疑降低了“洗稿”的难度,提高了侵权的风险。例如,混元助手、豆包等大模型具备生成短篇文章、改写文章的功能。大模型的强大之处在于可以在文字的基础上附带相关图片、数学模型等,这离不开大模型强大的信息收集和整合能力。

5. 视频内容的生成

AGI 大模型可结合图像、语音等元素,生成一段极为逼真的视频。例如,美国女歌星泰勒·斯威夫特一则说中文的短视频曾一度在各社交平台火了起来,有的播放量已经达到了 600 多万次,直到后来人们才发现这是 AI 生成的(机器之心,2023),这表明通用人工智能技术已经拥有了以假乱真的能力。例如,国外网站 HeyGen 可以选择人物的语言、性别和口音,并将人物粘贴到视频上,其所生成的视频内容经常被运用到游戏制作领域。

二、通用人工智能应用的著作权困境

(一) 通用人工智能"输入端"的著作权难题

通用人工智能大模型的培养和训练需要学习成千上万的数据样本,毫无疑问,这些样本将包含大量私有性质的版权作品。因此,在 2023 年全球首例控诉 AI 侵权案中,美国盖蒂图片社起诉 Stability AI,指控其"侵犯数百万图像版权"。随后,三位美国艺术家也加入了该阵营,控诉 Stability AI 从网络上采集了近 60 亿张图像来训练其 AI,侵犯了"数百万艺术家"的权利(wuhu 小精灵,2023)。本案被告以"并非完全匹配"为由进行抗辩,原告因证据不足而败诉。虽然没有取得诉讼的胜利,但法官表示艺术家们在收集到足够证据后,可以重新起诉。据《中华人民共和国著作权法》(以下简称《著作权法》),任何抄袭、修改他人作品的行为都是违法的。AI 所生成的图片是否仅仅只是抄袭和修改了相关作品,达到"实质性相似",以及是否构成合理使用,这是其是否构成侵权的关键。

1. "实质性相似"的认定困难

"实质性相似"常用于鉴定非文字作品是否构成侵权,前提是与疑似侵

权作品产生了接触。在我国,一般认为判断“实质性相似”的标准主要包括“整体观感法”和“抽象分离法”(梁志文,2015)。前者对比作品之间的直观感受,来确定是否构成“实质性相似”;后者将作品中不被《著作权法》保护的部分去除,从而在作品中受保护的部分之间做对比。通用人工智能时代下的 AI 绘画软件,其图片生成依靠的是其所学习、训练图片的风格特征向量,并没有对单一作品的全要素复制;并且由于 AI 模型对于上下文本理解的不确定性和随机性,其所生成的作品也具有一定的随机性。如果按照“整体观感法”,以 AI 模型对于类似风格的要素、布局的模仿程度,极有可能构成实质性相似。例如,AI 人物画中的赛博朋克风格,其背景以及色彩渲染、饱和度往往十分趋同,难以具有人类艺术家的创新点。但是按照“抽象分离法”,在 AI 模型生成随机性的支配下,难以达到认定侵权标准的“实质性相似”。相比之下,AI 音乐是否构成侵权的认定,也许更需要“整体观感法”的检验方式;出于音乐作品本身的开放性,依据“八小节”法,如果音乐旋律重复没有超过八小节,便不认定构成侵权。但是在通用人工智能时代,AI 模型绝对可以避免在八小节构成重复。因此,“整体观感法”十分重要,这种方法在乐评界也是检验作者创新能力和技术水平的重要方式。

2. 是否属于“合理使用”的范围

对于“合理使用”的范围,《著作权法》第二十四条规定,“为个人学习、研究或者欣赏,使用他人已经发表的作品”构成合理使用。在通用人工智能并不具备人格权的情形下,其综合学习他人作品并且生成相关的作品,无法满足“个人”学习和研究。同时,通用人工智能对他人作品的学习,也无法认为等同于其提供商作为“个人”进行的学习。但是,按照法律解释论,如果将通用人工智能大模型理解为“工具”,那么“个人学习研究”是否可以包含“个人利用工具进行学习研究”,从而进入“合理使

用”的范畴,为 AI“输入端”行为赋予合法性呢?同时,关于《著作权法》第二十四条“法律、行政法规规定的其他情形”,虽然其作为法律规范中常见的兜底条款,也许为某些情形提供了合理依据,但在整个通用人工智能模型“输入端”大环境下,无法为其训练、投喂模型的行为给予“合理使用”的正当依据。

(二)通用人工智能“输出端”的可版权性争议

美国版权局认为,艺术家克里斯·卡什塔诺娃使用 AI 绘画工具所生成的漫画书《黎明的查莉娅》(*Zarya of the Dawn*)的封面不应获得版权保护,此前获得的美国版权注册也应取消。这项裁决意味着,没有人类创作元素的 AI 生成的图像目前不能在美国获得版权(妙鼠品牌设计,2023)。在 2018 年“菲林诉百度”案中,法官认为数据报告是 AI 自动生成的,缺乏自然人的创作贡献,因此认定其不具有独创性,否认其具有版权。[①] 然而在 2019 年“腾讯诉盈讯科技”案中,对于腾讯公司主创团队使用 Dreamwriter 软件生成的新闻类文章,法院认为其体现了软件使用者(腾讯公司主创团队)独特的表达形式以及个性化安排,从而认为其具有独创性,视为法人作品并享有版权。[②] 可见在国内,对于人工智能生成物的态度以及司法实践是不确定的,偏向于根据具体情形进行判断。

《著作权法》第二条规定:“中国公民、法人或者非法人组织的作品,不论是否发表,依照本法享有著作权。”这意味着我国没有强调和明确作品创作主体的自然人身份,并且允许拟制人格的法人享有著作权。但是考虑到立法目的和历史背景,当时人工智能生成物还未进入法律实践,因此只有自然创作的“作品”才可以成为受《著作权法》保护的作品是理所应当的。并

① 北京互联网法院民事判决书(2018)京 0491 民初 239 号。

② 广东省深圳市南山区人民法院民事判决书(2019)粤 0305 民初 14010 号。

且我国并没有赋予通用人工智能拟制的法律人格,因此其无法享受著作权权益。洛克(1964)认为,将处于公共自然状态下的事物通过劳动转化为有价值的东西,付出劳动的人对其获得财产权,同时这种转化是有一定限度的,不能超出最大界限。劳动的主体只可能是自然人,AI 只是在程序员的算法设计下进行工作;并且 AI 的创作是一个“从有到有”的过程,并不是一个“从无到有”的过程。因此,不论从哪个角度看,依据洛克的劳动财产论对 AI 生成作品的行为进行解释,其都无法满足可以获得版权的条件。黑格尔(1979:59)《法哲学原理》关于人格权的相关论述认为,“人把他的意志体现于物内”,“必须对该物以外在的形式表现出来”,“必须便于他人承认”。概括而言,便是自由意志对于物的外在改造所产生的东西,其对此拥有所有权。但是,现阶段的 AI 模型所谓的“自主性”,很大程度上依赖人类的作用。“首先,人类需要对培训的数据素材进行选择,这决定了大模型的‘视野’;其次,需要人工对数据进行标注,从而进行价值顺位排序,这注定了大模型的‘喜好’。”(Ouyang, Wu & Jiang et al., 2022)而且,大模型思维模式不受情感、美感等人类感觉的支配,而是完全依靠数学逻辑的推演。AI 的自由意志在现阶段来看是不存在的,也不具备匹敌人类的创新能力。因此,无论是洛克的劳动财产论,还是黑格尔的人格权相关理论,都无法为 AI 生成作品的可版权性提供正当依据。总结来说,著作权法保护的客体是作品,而“著作权法意义上的作品是指人的思想或感情的独创性表达”(刘影,2017)。即便 AI 所创作的作品具备了人类的独创性,人类无法识别其属于自然人创作还是 AI 创作,但创作主体仍然不是人,因此在著作权法的意义上,其作品不受著作权法的保护。

但在实践中,情况是复杂交错的。“菲林诉百度”案与“腾讯诉盈讯科技”案属于人机协作的不同类型。“菲林诉百度”案中,原告使用 AI 生成了一篇数据报告,但文章属于菲林的独立创作。其创作的文章既符合洛克的

“出汗标准”,又符合黑格尔要求的自由意志表达,满足了著作权法意义上的独创性表达。同时,AI 所生成的报告属于 AI 大模型深度学习的结果,虽然满足独创性表达,但不符合自然人主体的消极要件。原告和 AI 共同的独创性表达造就了这篇文章,两者属于互相依存的合作关系。所以,在“菲林诉百度”案中,人工智能生成物属于人机合作作品,原告和 AI 属于合作关系。在“腾讯诉盈讯科技”案中,原告虽然利用了 Dreamwriter 撰写了文章,但其 AI 的本质是套用新闻模板的工具,其生成文章所依赖的数据、信息均来自原告,因此这份人工智能生成物只有原告付出了独创性表达,原告对此享有著作权。因此,在“腾讯诉盈讯科技”案中,人工智能生成物属于 AI 辅助作品,原告和 AI 属于工具辅助关系。但是问题在于,使用者使用 AI 生成的作品,仅仅 AI 做出了独创性表达,那么是否应该认可其作品版权,赋予哪方享有版权权益?

(三)通用人工智能生成物著作权归属问题

随着通用人工智能的发展,其生成的作品将会具有极大的变现能力。因此,对于通用人工智能生成物的财产权分配,以及人身权安排,是一个值得讨论的问题。AIVA 是 AIVA Technologies 打造的一款可以创作音乐的 AI 产品,可以通过阅读莫扎特、贝多芬、巴赫等人所创作的近 3 万首音乐作品来学习作曲。AIVA 已得到法国作曲家协会的资格认证,成为 AI 领域首款获得国际认证的虚拟作曲家(音乐财经,2018)。以 AIVA 所创作的专辑《艾娲》为例,虽然其所学习的音乐样本早已进入公共领域,但这个专辑并没有被投放于公共领域。AIVA 作为一个 AI 模型,享受了其生成作品的署名权。这个例子表明,通用人工智能可以拥有拟制的人格,从而拥有部分著作人身权。

关于通用人工智能生成物著作权归属的学说有多种。支持赋予其著作

权的观点认为,因为 AI 的生成内容具有作品的思想表现形式和人格主义要素,所以应受著作权保护,但其权利由参与创作或投资的自然人或法人行使(吴汉东,2020)。支持赋予其邻接权的观点认为,鉴于人工智能生成成果的保护价值与邻接权制度的价值相契合,因此可以赋予产生数据的程序或设备的使用权人享有对数据成果的数据处理者权(陶乾,2018)。不过邻接权的观点含义模糊,容易产生歧义,既可以从中推导出支持使用者的观点,也可以从中推导出支持投资者的观点。支持将其视为物权收益的“自然孳息”说认为,根据“天然孳息由所有权人取得”的法律规定,AI 是一种有形物,其生成物应归属 AI 的所有权人(叶霖,2019)。按照这种观点,应该将人工智能生成物的权利赋予 AI 的所有权人以及用益物权人。虽然人工智能大模型程序设计者在完成作品后会对其享有所有权,但在实践过程中,通用人工智能大模型往往是以公司职务作品的形式诞生的,因此这种学说实际上站在了投资人一方。“衍生”说认为,作品是人工智能程序的衍生作品时,程序员应该拥有版权;反之,作品是用户提供的某些基础作品的衍生作品时,用户或该基础作品的版权所有人拥有该版权(余祥、聂建强,2023)。此观点的亮点之处,在于关注了人工智能“输入端”版权者的利益。“公共领域”说认为,由于人工智能生成物不具有可版权性,因此应将其投入公共领域。著作权制度的重要价值在于平衡私人权益与公共利益之间的关系,将其投入公有领域供全社会自由使用,以保持公有领域思想材料的充分性,确保人类整体创新和创造能力的持续发展(徐家力,2023)。从卢梭社会契约论观点出发,著作权法是创作人和使用人之间的博弈,当创作人不是自然人时,理应将著作权归于使用人。但是,忽视人工智能生成物的权属性,将其投入公共领域,必然要承受其可能带来的负面后果。“拟制人格”学说认为,将 AI 本身视为作者,有利于厘清极其复杂创作过程中各方的利益关系,相关利益方的权益纠纷便会迎刃而解(杨延超,2019)。但是,该学

说缺少对人工智能的真实保护,只关注降低司法成本。“工具”说认为,只有人才能作为目的本身,人作为主体在与作为客体的人工智能交往过程中,主体人改造客体,并将其视为工具(李扬、李晓宇,2018)。此观点只适用于专用人工智能时代,而在通用人工智能时代,无法回应当今使用者“搭便车”和通用人工智能作品具备独创性这一著作权法上的积极要件的问题。

三、通用人工智能著作权困境的制度回应

通用人工智能所引发的著作权问题,暴露了现有法律规范体系的局限性。仅靠对现有法律语义的解释、立法目的的揣测,无法全然化解人工智能时代对著作权法律保护体系的挑战。诺内特、塞尔兹尼克(1994)为应对美国当时的社会问题,在《转变中的法律与社会》中构建了“回应型法”的制度图景。他们强调应走出法律规范因其维持内在合理性而独立于社会变动发展的封闭属性,而应以“目的”及不断变动的社会利益需求,作为法律制定、论证所围绕的重要方向,从而使得法律不断“回应”并适配社会需求,最终反作用于社会的繁荣进步。因此,为了回应通用人工智能的著作权困境,应发挥著作权法应有的社会功能,以平衡社会各方利益需求为目的,对通用人工智能著作权困境进行制度回应。

(一) 通用人工智能“输入端”的合法化辨析

数据是通用人工智能大模型的血液,如果没有这些数据进行学习,人工智能技术的发展将无从谈起。如果每项数据的投喂、学习均需要获得版权方的许可,即便版权方没有索要版权费,寻找版权方所消耗的人力、物力也将会使通用人工智能大模型技术的开发商承担巨额交易成本,从而不

利于通用人工智能技术的发展。高额的交易成本会降低协议达成的效率,并存在权利人利用劫持谈判策略的风险,因此资本的逐利性会驱使收集者最大限度地绕开壁垒,规避交易成本,造成侵权损害(郑飞、夏晨斌,2023)。

从功能主义视角看,著作权法是改善公共福祉作为社会功能的法律,对生成式人工智能的著作权问题进行探讨,还是要回归著作权法的规范逻辑(丁文杰,2023)。过于严苛、形式化地遵循著作权法律规范,可能会使著作权法的执行走向僵化,不利于实现其为创造财富、促进生产力发展的立法本意。《生成式人工智能服务管理暂行办法》(以下简称《暂行办法》)第六条规定:“推动生成式人工智能基础设施和公共训练数据资源平台建设。促进算力资源协同共享,提升算力资源利用效能。推动公共数据分类分级有序开放,扩展高质量的公共训练数据资源。鼓励采用安全可信的芯片、软件、工具、算力和数据资源。”这虽为我国人工智能“输入端”合法化建设指明了方向,但在同类型人工智能其智能水平竞争加剧的情形下,无偿的公共数据资源无法根本满足实践需求。对此,可以通过以下方式来化解。

1. 发挥版权集体管理、代理制度的功能

对于通用人工智能获取版权成本过大的问题,可以通过建立版权集体管理制度来化解。对于同意与人工智能生产方达成合作的版权方,可以集中并建立一个供训练人工智能的数据库。这个数据库中的作品,是默示授权给人工智能开发方的。这相当于一种法定许可,在法定许可模式下,人工智能在其生成作品的过程中使用他人在先作品不会构成侵权,但需要相关主体向作品权利人支付一定的使用对价(杨利华,2021)。对于作品授权方的利益补偿问题,可以通过事前与事后两种方式进行:事前方式是由版权集体管理方来统计各类版权使用费用,并由其集中向通用人工智能开发方收

取费用,随后进行分配;事后方式是通用人工智能在获得商业利益后,对于其生产所训练、投喂的版权作品库进行分红,并由版权集体代理方进行收取。

2. 设立针对 AGI 大模型训练的作品“禁用期”

《暂行办法》第七条规定,应“使用具有合法来源的数据和基础模型”,“涉及知识产权的,不得侵害他人依法享有的知识产权”。这虽然符合公平、正义等法律精神,但过于理想主义。首先,此要求所需的配套制度还未建立,因此对人工智能生产商的义务要求过多,不利于我国通用人工智能技术的国际竞争。其次,由于版权方与人工智能开发方在司法中并未实现证明责任倒置,因此举证开发方的训练数据侵权十分困难;再加上合法版权作品获得需要面对巨大的交易成本壁垒,从实用主义的角度看第七条,其无法产生实然的社会作用。因此,我们需要经过一个循序渐进的过程,来推动“输入端”的合法合规,这可以从对通用人工智能开发商对于数据、样本的收集进行一定程度的限制开始,如对作品建立区别于现有的版权制度,而针对 AGI 大模型训练限制较短时长的专项保护期。著作权法对作品的保护期限至少为 50 年,不符合知识更新迭代迅速的时代特征,不利于 AGI 大模型的发展,实现著作权法对作者和 AGI 开发商、社会公众的利益调和作用。建立针对 AGI 大模型训练较短时长的版权作品专项禁用期,不仅保护了人工智能的发展,还适度保护了自然人对于创新的贡献。

(二) 通用人工智能“输出端”的可版权性探索

通用人工智能大模型一旦建立,其内容输出的规模经济化将会使边际成本降至最低。通用人工智能大模型对市面大部分 PGC、UGC 将会产生巨大的代替作用,从而导致大部分自然人创作者在其行业领域被逐渐排除,进

而导致社会创新能力降低。自然人的作品是通用人工智能学习和进步的重要训练数据,自然人的创新动力不足会导致市面上创新型作品减少,进而使得通用人工智能大模型的训练数据质量下降。这将导致通用人工智能生成内容越来越同质化,因此只重视人工智能生成作品及其提供商的利益保护,反而不利于通用人工智能的进一步发展。如果坚决抵制 AIGC,将其作为侵权而产生的衍生物,抵制社会各界生成并利用的话,将会影响资源优化配置,阻碍社会生产力发展,并且不会激励自然人提升自身的创新能力。因此,对 AIGC 与自然人作品需要进行平衡保护。彻底否认 AIGC 的版权性不符合当今人机协作的时代背景,因此需要根据具体情况来进行考虑。

学界对于 AIGC 是否具有版权有不同的评判标准。有学者将“人类直接参与”作为评判标准,主张生成式人工智能有三种形式:完全由人工智能独立创作、由自然人辅助创作、按照自然人输入的提示词生成。上述三种形式中,只有“由自然人辅助创作”是人类直接参与的,这种情况下生成式人工智能生成的内容才可能具有版权(陈兵,2023)。这种标准的欠缺之处在于,忽视了自然人的独创性表达,将获取版权的标准降低到了只需要自然人的劳动。在很多情况下,使用者对于 AIGC 的贡献仅仅局限于对其修改、校对,并没有实质性地参与创作。另一种观点将“独创性表达的贡献者”作为标准,即 AIGC 作品中如果有自然人的独创性表达,那么便具有版权(丁文杰,2023)。这种观点虽然最符合著作权法的自然精神,但自然人不满足“最低限度的创造性”的劳动付出往往不构成独创性表达。将“人工智能发展阶段”作为标准(张金平,2023),在专家系统中由于人工智能仅起到工具作用,因此作品具有版权;下一等级的“一般互联系统”,由于人工智能不具备自主生成能力,只能在与自然人的合作关系中产生作品,因此具有版权;后期更高级的“高阶互联系统”,由于人工智能完全具备了自主生成能力,

因此不需要自然人的辅助,从而作品不具有版权。这种标准下的“高阶互联系统”无法解决人机合作的作品版权问题,因为有些 AIGC 需要自然人使用者与人工智能共同的创作元素,才能构成一个作品。例如,“菲林诉百度”案中的涉案文章,以及一首流行歌曲中人声、编曲、作曲的部分分别由自然人和人工智能合作完成的情况。依据独创性贡献赋予版权,符合理论层面的著作权法应有之义;但在现实层面仅仅考虑独创性贡献这一要素,无法在维持著作权法社会功能的前提下涵盖所有的利益相关主体。在“公平”与“效率”的语境中分析,独创性贡献仅能提供“效率”价值,无法兼顾“公平”,但“公平”是法律的重要原则和价值。“独创性贡献”与“工具辅助性贡献”在主体是人的情况下,其背后所付出的劳动可以有程度差别,而不应有性质差别,劳动没有高低之分。因此,有“工具辅助性贡献”的使用者对于劳动回报的利益需求,在难以寻求其他保护办法的情况下,需要赋予其版权予以保护。因此,结合前两种标准的“出汗原则”与“人类独创性原则”来界定是否有版权,是符合现阶段著作权法促进文化创作功能、平衡使用者劳动求偿需求的方案。

将人工智能自主生成物界定为无版权作品,具有合理性。著作权制度创立所依靠的激励理论,缺少对人工智能自主生成作品的激励必要。人工智能中计算机等高级机器属于机器范畴,没有独立的主体意识,无法享受作品创作带来的物质和精神利益。赋予机器排他权既无必要,也无实际意义,与激励理论激励创作、促进文化进步的初衷相悖(李扬、李晓宇,2018)。但是,在具备虚拟人格和作品独创性的情况下,AGI 应该获取版权,因为其满足了版权成立所需的积极和消极要件。

(三)通用人工智能生成物著作权归属的制度设想

通用人工智能技术的 AIGC 作品具有成本低、收益高等优势,将会产生

一块巨大的“利益蛋糕”。这块“蛋糕”怎么切,需要考虑许多目的。首先,需要考虑如何激励社会生产力的发展;其次,需要考虑社会不同群体的利益均衡,防止贫富差距过大;最后,需要考虑如何符合法治需求,有利于社会实质遵循。因此,有必要对 AIGC 作品按照人的参与情况进行分类考虑。

以 ChatGPT 为代表的通用人工智能带来了强人工智能时代,我们亟须赋予人工智能以拟制法律人格的理论和正当性支持。人工智能只要能够满足以下条件,就应该赋予其作者身份:人工智能生成物具有不可预测性且不是现有作品的重复;人工智能的创作行为是独立进行的,没有人类行为的干预和帮助;人工智能拥有是否生成作品和生成什么作品的决定权(朱梦云,2019)。我们可以参照《德国民法典》,确定虚拟“权利能力”作为“人”这一概念在实定法上的根据(朱振、都本有,2005),从而将通用人工智能视为具有权利外观的“形式人”,并赋予其一定的责任能力,使其拥有人的权利外观。因此,通用人工智能 AGI 只要提供了独创性贡献,便可拥有虚拟作者身份。在通用人工智能著作权的财产权划分上,可以将作品所得收益按照“自然孳息”说赋予其所有权人,即背后的投资人和人工智能“形式人”。由于公司一般为人工智能财产的所有者,因此其所得收益会分配给公司股东,最终受益者为投资人,因为投资人主体付出了技术性、经济性、组织性等方面的成本(许明月、谭玲,2018)。著作权中的人身权全部给予人工智能本身,其意义首先在于人工智能具备修改、撤回自身作品的能力,其作品不能被公司以及程序设计者更改。这是为了使人工智能独立创作作品保持完整性和特有价值,不会被其他人所扭曲。此外,这还意味着此类作品具备了署名能力,不会被相关自然人冒名顶替从而获得“不当得利”,违背市场经济道德。当然,通用人工智能享有这些权利的前提是公司设立了人工智能专项资金库,资金库的功能有两方面:第一,由于通用人工智能“形式人”具有

“权利外观”,因此需要具备承担侵权风险的能力,而且按照“形式人”能力大小其所需启动资金有高低标准,比如用户量大、社会关注度高的通用人工智能需要更高的注册资金;第二,通用人工智能由于具有权利外观,因此还需要具有收益能力,但人工智能并没有自然人的生理需求,其收益可以作为其运行、维护、维权以及升级迭代的资金。值得关注的是,通用人工智能在有权利外观的情况下也存在被侵权的问题,因为走法律程序维权也需要费用。

为何不直接将权利赋予人工智能的程序设计者,而将其收益给予公司及投资者呢?第一,程序设计者对于人工智能大模型本身享有著作权的激励,没必要对其进行双重激励。第二,通用人工智能作为强人工智能,具有自主学习能力与自主思维逻辑。编程者的预先程序设计与人工智能生成物之间并无必然的因果联系,前者更类似于人工智能生成物的创作工具而非创作来源(朱梦云,2019)。第三,程序设计者往往是公司的员工,员工在完成作品前后已经获得了公司的回报。第四,公司需要承担通用人工智能发展、运行和维护的各种成本。

1.“修改、检查”参与模式的 AIGC

在这种情形下,AGI 模型在 AIGC 作品中做出了独创性的贡献,而使用者仅仅付出了修改、检查的劳动成本。虽然其劳动付出在生产要素中并不具有稀缺性,但不可否认的是使用者付出了其他人没有付出的劳动。按照洛克的劳动财产论,劳动者有必要享受到“种瓜得瓜,种豆得豆”的自然规律,因此使用者有必要通过著作权、邻接权等方式享有一定的经济收入。不过,对于经济收入的取得方式,也可以通过其他方式获得其正当性。

假设赋予通用人工智能拟制的人格,让使用者与其建立一种“默示”的合同关系。使用者负责检查人工智能生成作品,通用人工智能大模型在作品获得收益后,按照合同约定对使用者进行经济回报。同样的思路,也可以

适用于使用者与人工智能大模型开发方的关系。这种以合同形式给予使用者以奖励的方式,不仅解决了使用者的劳动回报问题,符合自然法、劳动财产论的精神,同时可以使我国的著作权法逻辑自洽,解决了使用者在没有独创性贡献的情况下,依然需要获得经济鼓励的法律目的。因此,在使用者付出劳动的情况下,使用者有必要获取一定的财产权收入。

2. 合作贡献独创性的 AIGC

在“菲林诉百度”案中,人工智能与使用者共同的独创性成果催生了AIGC。在人机协作大背景下,此类 AIGC 作品只会越来越多。这包含两种情形:第一种情形是使用者对人工智能生成作品进行了重大的补充、修改和完善,也可以理解为一定程度上的“改编”;第二种情形是人工智能和使用者完成了同一作品的不同部分。“菲林诉百度”案就属于第二种情形,使用者在人工智能生成报告的基础上创作了一篇文章。

在各国的实践中,人工智能开发方都以会员制、付费制的模式来获取利润,AIGC 的最终收益由使用者享有。在没有付费的情况下,生成图片、音频等将会带有水印保护措施。通用人工智能网站或软件高额的生成费用,将会给使用者设立较高的准入门槛,而依靠 AIGC 作品盈利的少数人则获取了超出其付出的高额回报,这种情形不利于社会生产力的发展,以及社会团体间的利益均衡。而且,开发方在人工智能的创作中投入了劳动,人工智能的创作亦反映了设计人员的创造思想(党玺、王丽群,2021)。因此,有必要变革人工智能大模型开发商的经济激励模式,在 AIGC 作品在市场获取经济利益时,与其使用者共同分配利润,这样会使人工智能开发方拥有其他收益来源,降低“准入端”收费门槛,使更多用户加入人机协作创作大军,解放社会生产力,创造更多的社会财富。当然,这会导致使用者与开发方利益分割不清,加重司法机关负担,但其利益分割可以通过事前协议、事先约定的方式来完成,给予双方一定的自主性。在现实生活中,不同主体之间的利

益诉求复杂多样,而法律的应对能力有限,无法做到对所有可能发生的情形均予以合理规制,这时意思自治原则可以起到很好的补充作用(杨利华,2021)。在市场竞争中,拥有更加合理契约规则、更先进技术的人工智能大模型软件将胜出。因此,这种情形下著作权益不应只由 AGI 的使用者享有。

3. 人工智能独立创作的 AIGC

人工智能独立创作的 AIGC 包括人工智能自发创作的作品,也包括使用者要求人工智能做出,从而其被动创作的作品。如果将人工智能独立创作的作品投入公共领域,不承认其具有权利属性,将会引发许多问题。

卡什塔诺娃漫画书封面事件警醒我们,步入公共领域的 AIGC 作品可能会因为用户的不诚信行为而“搭便车”。人工智能生成作品在内容和表达形式上与人类作品高度相似,会出现自然人对 AIGC 作品冒名顶替的大量投机行为,从而产生“劣币驱逐良币”的不良风气,并助推“公地悲剧”现象的发生,不利于市场经济竞争秩序和社会创新务实精神的培养。

将人工智能独立生成的作品放入公共领域,将会导致“创新之源”的自然人创作者失去业务来源,从而失去经济支撑,不利于人工智能时代的社会团体利益平衡,以及通用人工智能技术的可持续发展。因此,反对将人工智能独立生成的 AIGC 投入公共领域,并支持其他方案的学说有许多。不管是赋予人工智能大模型开发商、投资方、设计人等对于 AIGC 自主生成作品的什么权利,都需要避免出现“搭便车”“冒名顶替”的不良行为,同时也需要在一定程度上维护自然人创作者的利益。

因此,人工智能独立创作的 AIGC 由于没有使用者直接参与,在其著作权权益分配上,可以将人身权赋予 AGI,财产权由公司与 AGI 共同享有。

四、结语

法律的合理性标准总是需要适配不同时期、不同地点社会秩序的图景，因此它也是法律的“理想成分”（庞德，2010）。以生成式人工智能为代表的通用人工智能技术极大地改变了我们当今社会的理想图景，因此现有的著作权法体系无力面对当前的社会利益格局，从而有走向失序之势。面对种种困扰，我们不能仅以单纯的注释和概念分析手段来应对，而是要以一种能动的“回应型法”解决方案，使法律恢复受损的社会功能系统，以促进社会最大福利为目的，来看待版权私有领域与公有领域的平衡问题、传统意义上概念语词的界定问题，以及新态势下既得利益的分配问题。

参考文献

陈抱阳，2023，《从生成式到多智能体的转向的人工智能艺术创作》，《美术研究》第 5 期。

陈兵，2023，《促进生成式人工智能规范发展的法治考量及实践架构——兼评〈生成式人工智能服务管理暂行办法〉相关条款》，《中国应用法学》第 4 期。

党玺、王丽群，2021，《人工智能生成物的著作权归属研究》，《浙江理工大学学报（社会科学版）》第 6 期。

丁文杰，2023，《通用人工智能视野下著作权法的逻辑回归——从“工具论”到“贡献论”》，《东方法学》第 5 期。

何大安、许一帆，2023，《人工智能应用扩张的经济学分析——兼谈 ChatGPT 对厂商经营活动的影响》，《社会科学战线》第 9 期。

黑格尔，1979，《法哲学原理》，范扬、张企泰译，北京：商务印书馆。

机器之心，2023，《口型几乎完美、还能卡点，霉霉说地道中文的视频火了，背后 AI 工具原来是它》，https://new.qq.com/rain/a/20231023A041IE00。

李扬、李晓宇，2018，《康德哲学视点下人工智能生成物的著作权问题探讨》，《法学杂志》第

9 期。

梁志文,2015,《版权法上实质性相似的判断》,《法学家》第 6 期。

刘影,2017,《人工智能生成物的著作权法保护初探》,《知识产权》第 9 期。

洛克,1964,《政府论》下篇,叶启芳等译,北京:商务印书馆。

妙鼠品牌设计,2023,《AI 绘画有版权吗? 全球首例官方判决来了!》,https://zhuanlan.zhihu.com/p/611096938。

诺内特、塞尔兹尼克,1994,《转变中的法律与社会》,张志铭译,北京:中国政法大学出版社。

庞德,2010,《通过法律的社会控制》,沈宗灵译,北京:商务印书馆。

陶乾,2018,《论著作权法对人工智能生成成果的保护——作为邻接权的数据处理者权之证立》,《法学》第 4 期。

wuhu 小精灵,2023,《全球首例大型 AIGC 侵权诉讼案件! 36 页诉状书状告 AI 侵犯图库老大哥数百万图像版权?》,https://mp.weixin.qq.com/s/DsPSs4ZrkoNhqOXqFOOEHQ。

吴汉东,2020,《人工智能生成作品的著作权法之问》,《中外法学》第 3 期。

徐家力,2023,《人工智能生成物的著作权归属》,《暨南学报(哲学社会科学版)》第 4 期。

许明月、谭玲,2018,《论人工智能创作物的邻接权保护——理论证成与制度安排》,《比较法研究》第 6 期。

《学术前沿》编者,2023,《通用人工智能的多维审视》,《人民论坛·学术前沿》第 14 期。

杨利华,2021,《人工智能生成物著作权问题探究》,《现代法学》第 4 期。

杨延超,2019,《机器人法:构建人类未来新秩序》,北京:法律出版社。

叶霖,2019,《人工智能生成物所涉著作权问题研究》,《科技与出版》第 5 期。

音乐财经,2018,《人工智能虚拟作曲家 AIVA 发布首张中国音乐专辑》,https://zx.ingping.com/c_1/51284.html。

余祥、聂建强,2023,《美国人工智能生成物的版权归属和可版权性标准的重构(英文)》,《科技与法律(中英文)》第 5 期。

张金平,2023,《论人工智能生成物可版权性及侵权责任承担》,《南京社会科学》第 10 期。

郑飞、夏晨斌,2023,《生成式人工智能的著作权困境与制度应对——以 ChatGPT 和文心一言为例》,《科技与法律(中英文)》第 5 期。

朱梦云,2019,《人工智能生成物的著作权归属制度设计》,《山东大学学报(哲学社会科学版)》第 1 期。

朱振、都本有,2005,《人格权的伦理分析》,《法制与社会发展》第 3 期。

Ouyang, L., J. Wu & X. Jiang et al. 2022, "Training Language Models to Follow Instructions with Human Feedback." https://doi.org/10.48550/arXiv.2203.02155.

人工智能生成技术方案可专利性界定路径*

宋红松　王瑞新**

摘要：如何界定人工智能生成技术方案的可专利性，是实务、司法和学界不容回避的问题。解决方法应回归对技术方案中自然人贡献的认定。挖掘人工智能生成技术方案成因，可以发现人工智能生成技术方案的本质是“自然人贡献+工具贡献”。自然人贡献不仅涵盖了自然人对技术方案的直接贡献，也包括对人工智能的构建、训练、调整等技术革新性贡献。工具贡献属于本领域普通技术人员的能力范围，随着人工智能技术的进步，本领域普通技术人员利用工具的能力水涨船高，使工具贡献本质上不具备“非显而易见性”。对待人工智能生成技术方案，应在现行专利法框架内合理确定技术方案中自然人的贡献，通过对比现有技术和本领域普通技术人员的能力，最终确定是否符合专利法的创造性要求。

关键词：人工智能生成技术方案　专利权　自然人贡献　非显而易见性　本领域普通技术人员

一、人工智能生成技术方案问题概述

人类推动科学技术进步的脚步总是会受到诸多因素的限制，而人工智能可以更快速、高效地生成、模拟和评估大量潜在解决方案。尤其是在复杂

* 本文系山东省重点研发计划（软科学）重大项目“山东省科技进步条例修订重点法律问题研究”（项目批准号：2022RZA01025）、烟台大学研究生科研创新基金资助重点项目“人工智能生成的技术方案的专利法定位”（项目批准号：KGIFYTU2403）的阶段性研究成果。

** 宋红松，烟台大学法学院。王瑞新，烟台大学法学院。

领域,人工智能能够有效解决很多复杂技术问题(Fraser, 2016)。自 2019 年开始,斯蒂芬·泰勒(S. Thaler)不断向德国、澳大利亚等几十个国家的知识产权部门申请一种名为“统一感知自主引导设备”(DABUS)的人工智能生成专利,并将人工智能填写为发明人,从而引发了人工智能生成技术方案的专利性问题。

DABUS 是一个自主人工智能系统,它通过人工神经网络分析和创造想法来产生发明。英国萨里大学教授阿伯特(R. Abbott)认为,社会需要人们使用人工智能来创造有益的发明,人工智能发明应该受到保护,以激励人们应用人工智能造福社会。在各个国家的判决中,澳大利亚联邦法院合议庭认为“授予专利的权利源于人类的努力”,并以 DABUS 并非自然人为由拒绝授予专利;德国专利局依据德国《专利法》第三十七条和《专利条例》第七条,以申请人必须提供发明人的姓名和地址等信息,以及获得专利申请权的路径和依据驳回了申请;中国国家知识产权局认为,专利法规定发明人享有的可以获得收益的财产权利以及署名以表明身份的人身权利,这些权利均属于民事权利,而民法规定享有民事权利的民事主体有三种类型,即自然人、法人和非法人组织,DABUS 是人工智能系统,不属于这三种类型中的任意一种,不能享有专利法上发明人相关的财产权利和人身权利,因此也无法在专利行政审批程序中被确定为发明人。

虽然各国在结果上驳回了 DABUS 的专利申请,但判决理由仅围绕人工智能是否符合发明人的定义,回避了人工智能生成技术方案是否具备可专利性这一核心问题。正如美国专利商标局《人工智能辅助发明的发明权指南》指出的:“尽管人工智能系统可能不会在专利或专利申请中被指定为发明人或联合发明人,但人工智能系统与其他工具一样,若其行为由人类执行,则可能构成发明权。”(Federal Register, 2024)

随着技术的进步,这一问题逐渐凸显出来。人工智能生成技术方案除

了应用于遗传编程、人工神经网络、机器人科学家等方面之外(Fraser, 2016),在结构生物学、医学领域的影响也逐渐增强。目前,开发新的有效药物需要耗费大量时间和资源,而人工智能可以在短时间内从数百万候选化合物中快速识别出潜在的生物活性化合物,从而加速新药的发现(Nayarisseri, Khandelwal & Tanwar et al., 2021)。2021 年,《自然》杂志发表了一篇运用人工智能工具 Alphafold 以高精度预测蛋白的文章(Jumper, Evans & Pretzel et al., 2021),而 AlphaFold 的设计者也因此获得了 2024 年诺贝尔化学奖。2023 年 3 月,英国科学、创新与技术部发布的《促进创新的人工智能监管方法》指出,人工智能技术可以预测已知几乎所有的蛋白质的结构,这一突破将加速科学研究和药物研发。

在 DABUS 案中,各国判决回避了人工智能生成技术方案是否具备可专利性这一问题,进而跳过了分析其中的自然人贡献是否满足了创造性这一核心步骤,直接通过发明人不符予以驳回申请。但如 AlphaFold 例中所见,哈萨比斯、詹珀以及泰勒等开发者、设计者在技术研发中发挥了重要作用,不仅其设计的人工智能生成物具有获得专利保护的潜力,其个人的贡献也直接影响了结果的产生。如何认定技术方案中自然人的突出贡献的性质,才是人工智能生成技术方案问题的核心。

目前,国内研究主要集中于探讨人类不介入的情况下人工智能生成技术方案的性质认定(刘瑛、何丹曦,2019),进而探讨申请过程中发明人问题的解决方案。其观点本质上认为技术方案是由人工智能生成的,不存在自然人贡献,因此并不能完全按照传统专利法思路解决这一类问题。但实际情况恰恰与此相反,“人工智能生成技术方案”这一概念具有极大的误导性,这里的“生成”是指输出而非创作,技术方案是由人工智能输出的,但技术方案的思路并不完全来源于人工智能,其中仍含有自然人贡献,恰恰是这部分自然人贡献导致人工智能生成技术方案可能获得专利保护。人工智能

生成技术方案可以良好地融入专利法体系,其本质为“自然人贡献+工具贡献”,通过确定其中自然人贡献是否达到创造性标准,进而决定是否赋予专利权。现有观点于解决实务问题并无裨益,问题核心在于如何正确分析人工智能技术方案中的自然人贡献。这也进一步产生了两个问题:第一,如何定性人工智能生成技术方案中的自然人贡献;第二,如何定性人工智能贡献的创造性。

二、人工智能生成技术方案中的自然人贡献定位

是否一定要以人工智能作为发明人?有学者认为,人工智能影响巨大,通过赋予人工智能以一定的法律主体地位可以解决诸多问题,同时也顺应了时代的发展(张惠彬、王怀宾,2024;季冬梅,2022)。也有学者认为,法律的主体应当限定为自然人,知识产权的保护目的便是围绕自然人所展开的,机器无法起到激励的作用(冯晓青、郝明英,2023)。站在传统观点来看,主体地位限定于自然人是知识产权保护的必然要求,其他主体的创造不属于知识产权所保护的范围。而站在现代观点来看,传统观点似乎无法解决发明人填写问题,且给予人工智能以一定的地位可以变相激励自然人。如果将人工智能作为发明人,前提是技术方案中不存在自然人贡献;而如果生成技术方案存在自然人贡献,则可以纳入现有专利法审查体系中。在现有体系下,可以通过分析自然人贡献进而确定是否赋予专利权。只有当结果完全不存在自然人贡献时,才有必要讨论人工智能发明人的问题。人工智能目前并非处于强人工智能阶段,其生成的技术方案中存在自然人贡献,现行专利法框架下完全可以容纳,因此没有必要将人工智能作为发明人,更不应当授予其主体地位。

技术方案的生成具备一个过程,最终的权利分配取决于过程中不同人

的贡献,做出创造性贡献的人将成为发明人。对于一项技术方案,应首先确定其中是否存在自然人贡献,再判断自然人贡献是否达到了创造性贡献标准,若符合则做出贡献的自然人成为发明人。而确定自然人贡献,应回到人工智能生成技术方案的过程。在这个过程中,人工智能生成技术方案存在自然人贡献,且自然人贡献是生成技术方案必不可少的一环。

首先,人工智能生成的技术方案往往是自然人和人工智能共同参与的。人工智能在其中被用于解决某些特定问题,包括人工智能在自然人具体指引下运算、自然人针对某些特定问题设计或修正人工智能程序的算法以符合其目的、针对人工智能训练语料进行筛选等。技术方案生成后,往往要经过自然人的修改和完善。因此,其中不可避免地存在自然人贡献,人工智能属于自然人的工具。

其次,即使是纯粹由人工智能生成的技术方案,其先天也存在自然人的参与。在人工智能生成技术方案的过程中,自然人的诸多行为在实质上做出了贡献,这些贡献是人类所特有的。自然人成功地将一种方案引入人类社会视角,以解决特定的技术问题。1474 年的《威尼斯专利法》规定,只要在“本城制造本城邦内先前未曾有人制造的”机械装置,就可以去登记并获得保护(刘春田,2022:160)。英国《垄断法》第六条中“真正的最初发明人”这一表述,在实践中曾被扩展到适用于未发明该发明但从国外引进该发明的人,认为引进发明的人应该得到奖励,因为他把以前为英国人所未知的发明引进了王国。在 2008 年修订以前,我国专利法“公开使用或者以其他方式为公众所知”的判别标准,仍然是相对新颖的。虽然这一方式逐渐退出舞台,但并不能否定其行为的价值性。可以类比经济学中“套利”与“创新”的概念:科技进步并非只有“创新”这一路径,引入发明也会起到关键作用。引入发明的认可与否是价值比较考量的结果,并非对行为价值性的拒绝,专利法承认这一行为的价值性。这一自然人贡献存在于所有技术方案中,具

有定性作用。人工智能生成数据自成为技术方案的那一刻起,先天必然经过人类特有的心智检验。某种人工智能生成技术方案能否解决技术问题,收到技术效果,需要人类心智的检验,这种检验实际上是人类发现技术方案的必经之路。这个过程包括人脑对可行性的逻辑判断、经验判断,以及利用客观事实去检验真伪的行为,以验证技术方案是否具备可重复性。人工智能生成技术方案在经自然人心智检验前,仍然处于不确定状态;只有心智检验符合逻辑或经验,或者利用实际操作验证具有可行性,才真正成为一种技术方案:我国专利法对偶然发明的认可也可以证明这一点。

这些贡献无法直接成为一项技术方案获得专利保护的理由,因此其存在性长期受到学界的忽视,但对于人工智能生成技术方案的定性分析具有重要作用。也正因此,人工智能生成技术方案中不可避免地存在自然人的参与。从定性上来看,人工智能生成技术方案存在自然人贡献。

实际上,技术方案中自然人贡献的范围要比我们通常认识的更为宽泛,对于结果的影响也并不仅浮于表面。2024 年 2 月 13 日,美国专利及商标局发布了人工智能贡献发明人资格的申请审查指南(Federal Register, 2024)。结论是,人工智能生成技术方案“并非绝对不可申请专利”,但发明人分析必须“关注人类的贡献”,这一贡献并不仅存在于表面。美国专利及商标局认为,人工智能系统的设计、构建或训练也可能属于对人工智能生成技术方案做出了重大贡献,针对特定问题或得出特定解决方案构建人工智能系统的自然人也可以是发明人。美国的指导意见并不排除自然人使用人工智能,只要该自然人对所要求保护的发明“做出了重大贡献”。这首先意味着人工智能生成技术方案含有自然人的“重大贡献”才是受到保护的理由,即将人工智能视为发明人构思发明的“工具”;同时,自然人“重大贡献”的范围不仅停留在其对结果直接的贡献上,也可以延伸到通过设计算法、训练系统等间接的贡献上。

有学者认为,人类至少在制作算法模型、选择和输入数据、编辑程序以指导计算机进行操作、向人工智能提出问题等一个或者多个方面做出了贡献(Kim & Horton, 2020),甚至包括认识到生成物的性质和用途而使该生成物成为发明(Burk, 2021)。在指南中,美国专利及商标局指出自然人的突出贡献需要满足下列标准:1. 以某种重要方式对发明的构思或付诸实践做出贡献;2. 对权利要求中的发明做出贡献,且该贡献与完整发明的质量相比并非微不足道;3. 所做的贡献不仅仅是向真正的发明人解释众所周知的概念和/或当前的技术状态。虽然人工智能生成技术方案本质上是“自然人贡献+工具贡献”,但是部分自然人贡献直接融入了人工智能本身的进步当中。实际上,能达到突出的贡献程度的“工具贡献”应当属于自然人贡献,因为正是自然人对算法和系统的贡献导致该工具较一般工具具有优势,进而做出突出的贡献。这些自然人贡献常隐藏在人工智能的外表下,跟随人工智能输出而呈现出来。美国关于自然人的突出贡献的标准没有触及创造性问题的本质,但也意识到自然人贡献的地位至关重要,并将之达到某种程度便作为人工智能生成技术方案获得保护的关键,实质上指向了自然人贡献的复杂性——直接贡献和借助人工智能工具的贡献。

也正因此,即使人工智能获得了权利,真正行使权利的仍是背后的自然人。人工智能生成技术方案的行为与人工智能背后的研发行为无法区分,人工智能的突出贡献实际上是自然人贡献。同时,人工智能本身属于人工智能开发者的个人财产,人工智能专利权也因此将会被视为个人专利权,最终也会被定性为开发者的个人权利。开发者作为人工智能及其财产(专利权)的所有人,实质上仍然是控制人工智能系统的人控制人工智能的权利。DABUS 专利申请背后的组织也承认这一事实(Bultman, 2021),他强调该系统的开发者应该是相关专利权的所有者。澳大利亚联邦法院承认 DABUS 为发明人的判决也支持了这一观点,该案的主审法官指出,泰勒显然

有权拥有发明,因为他是 DABUS 的所有人、程序员和操作员(Sun, 2022)。同时,DABUS 诉讼案表明,人工智能开发者泰勒可能会将 DABUS 的商业活动和他自己的商业活动混为一谈。泰勒可以向他人授权、发放许可甚至转让以收取费用,他的行为将与 DABUS 自身的商业活动产生冲突(Sun, 2022)。人工智能并不能行使权利,也无法获得激励,本质上只是自然人在操控人工智能,最终权利的决策仍然属于自然人。因此,应直接将对人工智能及其技术方案做出贡献的自然人作为法律规制的主体,而不能将人工智能作为承担行为后果的法律主体。

人工智能生成技术方案中存在自然人贡献,而自然人贡献是发明人问题关注的重点。人工智能生成技术方案的本质是“自然人贡献+工具贡献”,其中能满足突出贡献标准的只能是自然人贡献。若人工智能生成技术方案能够满足授权标准,那么自然人贡献也必然满足突出贡献标准,此时应将做出该贡献的自然人直接填写为发明人。同时,针对自然人贡献给予肯定和保护,将直接激励自然人对技术的探索和公开,促进科学技术的进步。专利法的目的是规范自然人的行为,人工智能并没有发展到所谓“强人工智能”时代;即使这一时代真的到来,法律所规范的仍然是站在人工智能背后的自然人行为,人工智能在专利法面前仍应作为工具。专利法保护真正的自然人贡献,目前并不需要以人工智能作为发明人来解决问题,发明人的判别标准仍是做出创造性贡献的自然人。

人工智能生成技术方案中发明人问题的深层原因,在于人工智能工具贡献的性质模糊。人工智能工具贡献确实远超预期,其计算处理能力使得诸多复杂问题迎刃而解,其生成与输出能力让技术方案的发明人难以确认,这是许多学者动摇发明人标准的核心诱因。人工智能生成技术方案的创新点往往混合了工具贡献和自然人贡献,而对工具贡献的定性是判断创造性的前提。人工智能工具贡献若具有创造性,学者主张在专利法上予以特殊

地位无可非议;但人工智能工具贡献并不具备创造性,其本质仍属于本领域普通技术人员的能力范围。

三、人工智能生成技术方案的创造性定位

(一) 人工智能生成技术方案的本领域普通技术人员定位问题

技术方案中的人工智能工具贡献是否具有创造性?如果人工智能生成技术方案中的人工智能工具贡献和自然人贡献两者混同了,将如何区分(李青文,2023)?创造性的判断应将本领域普通技术人员作为参照系。本领域普通技术人员的定性将直接影响专利法对人工智能生成技术方案的态度,本领域普通技术人员利用人工智能工具的能力也直接影响人工智能工具贡献的性质。在人工智能领域中,对本领域普通技术人员定位是解决创造性问题的关键。

本领域普通技术人员是一种假设的“人”,其知道申请日或优先权日之前发明所属技术领域所有普通技术知识,能够获知现有技术,并且具有应用该日期之前常规实验手段的能力。本领域普通技术人员的知识量并非一成不变的,其掌握的技能也随技术发展而拓宽,而人工智能也会逐步进入本领域普通技术人员的视野。本领域普通技术人员不会对人工智能技术所带来的便捷熟视无睹,反而会恰当接受这一解决问题的工具,他会逐步学会一定人工智能技术,掌握应用人工智能解决问题的一定能力,他的能力范围决定了人工智能工具贡献的性质。

随着人工智能技术的发展,本领域普通技术人员的标准也需要相应提高,合理对应本领域普通技术人员对人工智能的熟练程度,而这种熟练程度实际上进入了人工智能工具贡献的范围。利用人工智能的优势解决问题,

本属于本领域普通技术人员可以想到的解决方案。人工智能工具能够做出贡献,但若能做出突出的贡献,其相较于普通工具必然具有改进,这部分构建或改进实质上属于自然人贡献的范围,是自然人实质上对结果做出贡献。换言之,人工智能技术不断发展,本领域普通技术人员应用人工智能的能力水涨船高,使得人工智能工具贡献实际上属于本领域普通技术人员的能力范围,其贡献的创新点缺乏“非显而易见性”,进而无法满足创造性要求。

人工智能工具贡献无必要进行保护,以自然人贡献和本领域普通技术人员作为判断的核心才具备价值性。“工具贡献”属于本领域普通技术人员应用必备工具就能实现的范围,其水平随着人工智能的发展而提升,其贡献不具有“非显而易见性”,对工具贡献进行保护只会阻碍技术的发展,成为巨大的“专利丛林”。只有对普通技术人员不曾掌握的自然人突出贡献进行保护才具有专利法意义,才能真正起到激励自然人、促进科技进步的作用。

对于两者混同或难以判断来源的创新点,仍应将本领域普通技术人员作为标准,本领域普通技术人员使用人工智能可以解决的问题不具备创造性。即使发明人并没有利用人工智能,也应按照上述审查流程认定为不具备创造性。这样即可以避免“专利丛林”的形成,促进人工智能的应用发展;同时通过此种认定,在申请程序中也可以起到类似“证明责任倒置”的效果,申请人将着重证明其存在自然人独特的贡献,撇清与人工智能混同的嫌疑,使专利审查员可以有效应对人工智能时代的专利冲击,有利于审查的进行,同时促进技术方案的充分公开。

(二) 人工智能生成技术方案的创造性判断

在专利法的三种判断中,学界对于人工智能生成技术方案的实用性和新颖性问题并无很大争议。人工智能生成技术方案可以解决一些之前无法

解决的技术问题(吴汉东,2019),而且将生成技术方案与全部现有技术相对比,尽可能使该技术方案符合新颖性要求,这对于人工智能来说并不困难(杨利华,2023)。但是,人工智能生成技术方案是否具有创造性?从人工智能作品的角度看,有学者认为人工智能只能按照人类预先设定的算法、规则和模板进行计算并生成内容,其结果具有高度的可重复性,从而不符合独创性的要求(王迁,2017)。自然人使用预先的算法、规则和模板所得到的有限的或者排列方式不同的组合发明或选择发明,是否符合创造性要求?

在 1952 年废除"天才闪光理论"时,美国国会得出专利制度应漠视发明产生手段的结论。"天才闪光理论"是一种无益的理论,因为它含糊不清,下级法院难以解释,涉及法官对专利权人思想状态的主观判断,并且大大增加了专利申请的难度。1943 年,美国国家专利规划委员会就"天才闪光理论"提交报告,认为专利性应客观地根据对技术进步所做贡献的性质来确定,而不是主观地根据完成发明的过程的性质来确定。美国 1952 年《专利法》采纳了这一建议,在立法上放弃了"天才闪光理论",并在第一百零三条中明确规定,"可专利性不因发明被制造出来的方式而受影响"。美国最高法院在"格雷厄姆诉约翰迪尔公司"案中提出诱导标准,认为专利法的"非显而易见性"原则旨在将专利授予仅限于"那些除非诱导专利而不会被公开或设计的发明"①。虽然人工智能的算力基础是排列方式或者算法,但当其应用在诸多未曾利用人工智能解决问题的领域时,完全可以利用其优势解决一些长期困扰自然人的技术问题。因此,即使是有限的排列组合,如果能够取得意料不到的技术效果,则其仍具备突出的实质性特点和显著的进步,可以具备创造性,并不违反专利法的要求。

① Custom Accessories, Inc., Plaintiff-appellant, v. Jeffrey-allan Industries, Inc., Defendant-appellee, 807 F. 2d 955 (Fed. Cir. 1986).

然而,创造性的判断应回到自然人贡献的标准当中。创造性判断是一个基于现有技术,通过对比技术方案与现有技术确定一项技术方案是否具备突出的实质性特点以及显著的进步的过程。因而,随着人工智能技术的发展,现有技术的内容和范围不断增加,人工智能生成技术方案的“非显而易见性”判断标准也呈现出一种变化,即随着本领域普通技术人员标准的提高,人工智能生成技术方案的“非显而易见性”标准也在提高。在对创新点进行创造性审查时,首先应区分自然人贡献与人工智能贡献(朱雪忠、张广伟,2018),以本领域普通技术人员为标准,利用人工智能生成技术方案贡献公式“自然人贡献+工具贡献”分别进行判断。如上文所言,“工具贡献”如果表面能满足创造性突出贡献的标准,实质上应当属于“自然人贡献”,因为自然人贡献可以延伸到设计算法等方面,而正是算法或系统革新才导致工具相较于先前能产生突出的贡献,“工具贡献”应属于本领域普通技术人员的能力范围,贡献部分始终不具有“非显而易见性”。“自然人贡献”部分需要对比本领域普通技术人员能力范围,确定是否具有“非显而易见性”,进而决定是否具有创造性。

创造性应围绕本领域普通技术人员判断。目前,人工智能技术尚未成熟,率先更新算法及系统,利用其特有的算力优势解决本领域普通技术人员不容易解决的问题,这部分贡献属“自然人贡献”,其突出贡献超过了本领域普通技术人员的能力水平,得到的技术方案可以具备创造性。甚至在人工智能尚未应用于某一领域、本领域普通技术人员想不到使用人工智能来解决问题时,率先使用人工智能这一工具解决问题产生的技术方案也可能具备创造性,这本质上是自然人引入人工智能这一工具,贡献具有达到标准的可能性。只因诸多领域中普通技术人员逐渐利用人工智能解决问题,人工智能成为普通技术知识和常规实验手段,成为很容易想到的问题解决工具,标准进而提高,导致贡献不满足创造性要求。

此外,人工智能代码开源会影响人工智能生成技术方案的可专利性。人工智能若将其代码进行开源,将成为本领域普通技术人员可以获得的工具,其生成的技术方案也可能随之失去新颖性和创造性。因此,如果对人工智能生成的技术方案进行保护,可能不利于鼓励代码开源,反而促使企业转而使用商业秘密进行保护。而在人工智能成为本领域普通技术人员的技能之前,否认率先解决技术问题的技术方案的创造性并不合适,过早否定其创造性将不利于科技和产业发展。但广泛应用后仍然进行专利保护不仅不利于鼓励代码开源,且易形成“专利丛林”。这需要综合各项因素,准确把握人工智能在行业中的发展程度,合理判断本领域普通技术人员能力标准。专利管理部门以及司法机关在个案中应结合不同行业、不同领域人工智能发展的现状及需求,灵活合理地确定本领域普通技术人员对人工智能的掌握程度,调和代码开源与专利保护的矛盾,避免形成“专利丛林”,探寻激励人工智能技术进步和产业发展的最优路径。

四、结论

人工智能生成技术方案只有其自然人贡献满足创造性审查标准,才能获得专利保护。人工智能在科技创新领域的影响,已经成为专利法不容忽视的挑战(尹志锋、曹爱家、郭家宝等,2023;陈伟、邓堯、杨柏等,2024)。对人工智能生成技术方案进行自然人贡献的分析、创造性问题的认识、普通技术人员标准的确定,将直接影响其在专利法中的地位,进而影响行业科技发展模式的选择。对待人工智能生成技术方案,不应仅停留于“三性分析”的表面,更应深入探寻其生成过程。在这个过程中,自然人贡献始终存在,人工智能处于工具地位,自然人贡献决定了技术方案是否具有创造性。现有

观点中的发明人填写论、实质贡献论、间接激励论都站不住脚。自然人贡献始终存在且占据核心地位,发明人并不需要填写为人工智能;也不需要给予人工智能贡献以特殊地位,工具贡献实质上不具有“非显而易见性”,不存在突出的贡献;直接对自然人激励能起到更好的效果,也不需要通过人工智能进行间接激励。正因自然人贡献具有突出的地位,其能否满足突出贡献标准决定了人工智能生成技术方案是否为专利保护的对象,做出突出贡献的自然人当然可以成为发明人,因此发明人问题完全可以在现有体制下解决。审查过程需要重点关注本领域普通技术人员,综合多种因素合理确定能力范围。人工智能的普及会使本领域普通技术人员的技术水平水涨船高,人工智能成为必备工具,使得工具贡献本质上属于本领域普通技术人员利用合理工具能达到的范围,不具有“非显而易见性”。合理确定本领域普通技术人员的能力范围可以协调好保护与发展的冲突,避免专利保护阻碍技术和经济发展。对待人工智能生成技术方案,应先找到背后的自然人贡献,并利用“自然人贡献+工具贡献”的判别标准合理确定其实质范围,再通过本领域普通技术人员标准判断其是否具备创造性。只有正确认定并保护真正自然人的突出贡献,才能使得专利法正确定性并接纳人工智能生成技术方案。

参考文献

陈伟、邓堯、杨柏等,2024,《人工智能能否赋能区域技术创新——创新环境的中介效应》,《科技进步与对策》第 10 期。

冯晓青、郝明英,2023,《人工智能生成发明专利保护制度研究》,《湖南大学学报(社会科学版)》第 2 期。

季冬梅,2022,《人工智能生成发明专利保护的理论回归与制度探索——以 DABUS 案为例》,《科技进步与对策》第 18 期。

李青文,2023,《人工智能对发明人制度的挑战及应对策略》,《中国科技论坛》第 8 期。

刘春田,2022,《知识产权法》,北京:中国人民大学出版社。

刘瑛、何丹曦,2019,《论人工智能生成物的可专利性》,《科技与法律》第 4 期。

王迁,2017,《论人工智能生成的内容在著作权法中的定性》,《法律科学(西北政法大学学报)》第 5 期。

吴汉东,2019,《人工智能生成发明的专利法之问》,《当代法学》第 4 期。

杨利华,2023,《人工智能生成技术方案的可专利性及其制度因应》,《中外法学》第 2 期。

尹志锋、曹爱家、郭家宝等,2023,《基于专利数据的人工智能就业效应研究——来自中关村企业的微观证据》,《中国工业经济》第 3 期。

张惠彬、王怀宾,2024,《生成式人工智能时代发明人制度的结构性改革》,《北京邮电大学学报(社会科学版)》第 4 期。

朱雪忠、张广伟,2018,《人工智能产生的技术成果可专利性及其权利归属研究》,《情报杂志》第 2 期。

Bultman, M. 2021, "Can a Robot Invent? The Fight Around AI and Patents Explained." https://news. bloomberglaw. com/ip-law/can-a-robot-invent-the-fight-around-ai-and-patents-explained.

Burk, D. 2021, "AI Patents and the Self-Assembling Machine." *Minnesota Law Review Headnotes* 105(2).

Federal Register 2024, "Inventorship Guidance for AI-Assisted Inventions." https://www. govinfo. gov/content/pkg/FR-2024-02-13/pdf/2024-02623. pdf.

Fraser, E. 2016, "Computers as Inventors: Legal and Policy Implications of Artificial Intelligence on Patent Law." *SCRIPT-ed* 13(3).

Jumper, J., R. Evans & A. Pretzel et al. 2021, "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596.

Kim, A. & M. Horton 2020, "Inventorship: Why AI is not Smart Enough Yet." *Managing Intellectual Property* 286.

Nayarisseri, A., R. Khandelwal & P. Tanwar et al. 2021, "Artificial Intelligence, Big Data

and Machine Learning Approaches in Precision Medicine & Drug Discovery." *Current Drug Targets* 22(6).

Sun, H. 2022, "Artificial Intelligence Inventions Artificial Intelligence Inventions." *Florida State University Law Review* 50(1).

上海数字适老 2.0 探索*

——构建新旧媒介并行的微粒社会

张　林**

摘要：在数字化浪潮里，老年群体的数字融入问题愈发凸显。本文把老年群体分为三类，即数字技术熟练使用者、“数字移民”和“数字难民”，并基于全球在地化视角，探讨上海在数字技术适老化 2.0 升级行动中，怎样构建老年友好型数字包容下新旧媒介并行的微粒社会。经由实地调研新加坡、瑞士、日本等国的经验，剖析其相关做法与实践案例，提出上海的具体实施路径。本文指出，培育老年友好型数字包容、保留传统服务方式、推进科技无感化转换、构建全方位立体化的支持网络，以及激发老年群体社会参与和数字消费的内生动力等，可有效提升老年人的数字素养，推动老年群体的数字融入。

关键词：数字适老　银色数字鸿沟　数字移民　数字难民　微粒社会

一、引言

在数字化时代，老年群体面临的数字鸿沟日益显著。这一问题不仅影响老年人的生活质量，还制约其社会参与度。近年来，尽管政府、社会、企业等多方面在数字适老化政策研究、电信服务、移动互联网应用等方面付出诸

* 本文系 2023 年度上海市人民政府决策咨询上海大学研究基地课题“上海推进养老服务高质量发展研究”（项目批准号：2023-JD-P09）的阶段性研究成果。

** 张林，上海大学新闻传播学院、上海大学全球人工智能媒体传播研究院。

多努力,但老年群体的数字融入问题仍未从根本上得到解决。尤其在数字技术快速发展的背景下,老年人在新旧服务方式衔接时面临诸多挑战,如技术接受度低、操作能力不足、对新技术感到焦虑以及依赖传统服务方式等。这些问题致使老年群体融入数字社会的进程缓慢,数字鸿沟依旧存在。

为应对这一挑战,本文从全球在地化视角出发,探讨上海在数字技术适老化 2.0 升级行动里,怎样构建新旧媒介并行的微粒社会,从而推动老年人深度融入数字环境。本文借助实地调研新加坡、瑞士、日本等国经验的方式,对它们在技术包容和心理包容双重维度下的成功实践予以分析,并给出上海的具体实施路径。本文指出,保留传统服务方式、推进科技无感化转换、构建全方位支持网络、激发老年群体内生动力,能够有效提高老年人的数字素养和生活质量,构建老年友好型数字包容社会。

二、文献回顾与研究问题

(一) 老年数字鸿沟:并非仅存“有无”之分

为推进老年人生活的“数字无障碍”,近年来,政府、社会、企业、家庭和个人五位一体,在数字适老化政策研究、电信服务、移动互联网应用、硬件终端、标准规范制定与应用以及终端适老化评测等多方面开展了信息无障碍的适老助老工作,推动老年群体融入数字社会,畅享数字美好生活。工业和信息化部称,截至 2024 年 1 月,全国已有 2577 家网站和 APP 完成适老化及无障碍改造(新华社,2024),下一步即将开展数字技术适老化 2.0 升级行动。

在数字技术的赋能下,我国每两名老年人中就有一人接触互联网。截至 2024 年 6 月,我国 60 周岁及以上老年人口近 2.97 亿人,其中网民达

1.5725 亿人,占全国网民总数的 14.3%,较 2023 年 12 月新增 742 万人,其中 60 岁及以上群体新增占比为 20.8%(中国互联网络信息中心,2024a)。数字化生活已经成为老年人的一种生活范式。

但与此同时,有一点容易被忽视:一方面,已踏上数字化快车的老年群体仍然面临“科技恐惧”与“技术困境”,多数老年网民作为“数字移民”(digital immigrants),无法独立完成常见的网络活动(李清华,2022);另一方面,截至 2024 年 6 月,我国还有上亿老年人未能踏上数字化快车,被“隔离”在数字化生活之外,在非网民总数中的占比高达 62%(中国互联网络信息中心,2024a),成为“数字难民”(digital refugees)(Prensky, 2012),而 2023 年 12 月这一比例仅为 39.8%(中国互联网络信息中心,2024b)。这表明,近两年我国老年群体“银色数字鸿沟”(grey digital divide)的弥合速度低于数字化速度,不少老年人没能跟上数字社会发展的节奏。

这意味着,此前学者基于代际语境(青年一代 VS 老年一代),对信息社会整体进行“有无”数字鸿沟区分的二元信息分层方式(Prensky, 2012: 88;陆杰华、韦晓丹,2021),也就是认为作为“数字原住民”(digital natives)的青年一代没有数字鸿沟,而老年一代存在数字鸿沟的分层方式,在面对老年社会语境时,需要基于老年群体内部语境对这一群体进一步更新并细分为三元信息分层。参考学者邱林川的研究实践(邱林川,2019),笔者认为老年社会的三元信息分层应为:“信息拥有者”(haves),即数字技术熟练使用者;“信息未拥有者”(have-nots),即“数字难民”;还有大量的“信息中下阶层”(information have-less),即在信息社会分层结构里介于前两者之间的人群,如已经搭上数字快车但仍存在技术障碍、无法独立完成网络活动的老年“数字移民”。

基于老年社会内部更新后的信息分层结构,笔者认为:新技术的发展不会因老年人而放慢脚步,弥合“银色数字鸿沟”是个漫长的过程。在未来开

展数字技术适老化 2.0 升级行动时,相关部门和实施主体一方面可积极利用大数据、人工智能大模型等创新技术,助力实现敏捷、精准、高效的服务;另一方面应充分考虑数字技术的补偿性或替代性手段(卜卫、任娟,2020),保留一些“老方式”,平衡好传统与现代的关系,构建新媒介与传统媒介并存的老年友好型数字包容社会,给予老年人更多的选择。

(二) 数字适老化改造:同步保留一些“老方式”

在国内外,数字适老化改造中保留一些“老方式”这一议题,已被纳入政策与媒介议程。2020 年 10 月,欧洲理事会专门讨论数字时代的老年福利,呼吁加强代际融合,确保加速发展的数字化(尤其是健康、社会服务、养老服务等领域)给老年人带来便利,同时为他们保留非数字化服务选项(张朋辉,2021)。同年,国务院办公厅印发《关于切实解决老年人运用智能技术困难实施方案的通知》,要求在各类日常生活场景中保留老年人熟悉的传统服务方式,以充分保障运用智能技术困难的老年人的基本需求(国务院办公厅,2020)。国内主流媒体与民众对保留“老方式”的呼声颇高,如《人民日报》发表多篇文章强调,推进数字化时必须保留传统服务方式,确保老年人顺利办理各项业务;《光明日报》呼吁社会各界关注老年人数字转型中的困境,提供更多线下服务与支持。在线调查显示,超 70%的受访者支持数字化服务中保留传统服务方式,认为这有助于老年人更好地融入社会(王志伟,2021)。

但在实际操作中,进行数字适老化改造时,如何保留部分“老方式”呢?梳理国内外文献可知,国内外学者重点研究了该问题的前半部分,也就是各个国家怎样弥合“银色数字鸿沟”,优化数字适老化体验。例如,欠发达国家和地区侧重于基础设施的补充与建设,发展中国家着重推动数字化转型和设备普及,发达国家则秉持积极老龄化(active aging)理念促使数字包容

(digital inclusion)、适老化(elderly-oriented)和智慧养老(smart senior care)等理念在数字社会中得以实现(周煜,2022)。然而,针对数字适老化改造时保留部分“老方式”的具体措施与路径,中外文献几乎未曾涉及,学界迫切需要加快开展理论与对策研究。

尽管如此,日本、新加坡、瑞士等深度老龄化国家,已在这一议题上积累了一定的实践经验(Quan-Haase, Mo & Wellman, 2017)。日本率先进入超级老龄化社会,一方面对智能化设备按照日常生活细节进行改造,以降低老年人的学习成本和认知负担;另一方面并没有缩减或取消原有的人工服务系统。新加坡是亚洲老龄化速度最快的国家,该国推进数字适老的协同治理举措,在确保公共服务可及性的同时,保留了非数字化服务方式的选择,从而让老年人获得更多的情感慰藉以及对数字社会的融入感。瑞士在不少人眼中是“最适合养老的国家”,也是全球老龄化较严重的国家,该国在为每位老年人建立“云”账户的同时,着力加强邻里互助,提升公共空间对老年人的便利性,打造老年友好型社会。

(三)上海“全球在地化”探索:构建新旧媒介并行的微粒社会

上海是全国老龄化程度最高的超大城市之一,既具有全球化特点,如老年人口总量大、增速快,老年人对精神慰藉和情感陪伴需求度高(张虎祥,2023);又有在地化特色,如城市数字化与人口老龄化进程相互交织、高度叠加,人口结构的“老”与媒介技术的“新”产生强烈碰撞,数字技术基础设施和智能化的高速发展与人口老龄化程度持续加深之间的矛盾加剧(方提、张潇文,2023),老龄化与高龄化、失能化、空巢化、独居化、少子化“五化”并发,老年群体内部数字化态度多元、水平参差不齐、需求异质具体。

就数字适老化改造与保留传统服务方式的关系而言,上海市“十四五”规划纲要提出,要着力提升智能化服务的适老性水平,完善日常生活中传统

服务方式的兜底保障,推动老年人共享便捷智能的数字生活(上海市人民政府,2021)。《上海市全面推进城市数字化转型“十四五”规划》指出,要坚持效率与温度兼容并蓄,统筹处理好公平与效率的关系,既要通过数字化转型提升城市系统的效率,又要以市民的实际感受为考量,关注“数字鸿沟”问题,构建线上线下有机联动的数字孪生生活空间(上海市人民政府办公厅,2021)。然而,在实际推进过程中,如何实现线上线下有机联动?如何统筹处理好公平与效率的关系?这些具体问题都有待在未来开展数字技术适老化 2.0 升级行动时解决。

笔者认为,对于上海而言,扎根地方实际,借鉴日本、新加坡、瑞士等国家经验,把全球化和在地化相结合,基于全球在地化(glocalization)视角(Robertson, 1995),统筹数字适老化改造与保留传统服务方式的关系,构建新旧媒介并行的微粒社会,是一种可行的探索与创新。

“微粒社会”(granular society),也称“精细化社会”,是德国社会学家克里斯托夫·库克里克在《微粒社会》一书中提出的社会学概念。它是对数字化时代一种新型社会形态的形象化描述,是一个颗粒度微小且精细的社会(库克里克,2018:66)。在微粒社会中,社会的数字化程度达到了前所未有的深度。个体之间细微的差异被更加深入地解析,这些差异也随之被放大,个体从而变成“微粒人”。对群体需求的同质化满足,将转变为对高度差异化的个体需求的满足,进而转变为对同一个体在不同时间、空间以及生命历程中需求的满足。在这种社会形态下,大众的需求能够以更加微观、具体的方式呈现出来,这有利于政府部门和社会组织更好地提供多样、分层、有针对性的服务,从而更好地开展社会治理。

构建新旧媒介并行的微粒社会,需精准锚定不同信息阶层老年人的信息需求,统筹好数字适老化改造与保留传统服务方式的关系,把对老年群体数字需求的普惠性满足转变为对不同信息阶层老年个体需求的精确满足,

进而转变为对同一老年个体在不同时间、场景和生命历程下需求的满足,为其提供多层次、个性化的数字服务。

因此,本文的研究问题是解析国际社会统筹数字适老化改造与保留传统服务方式关系的相关做法和经验,探讨上海构建老年友好型数字包容下新旧媒介并行的微粒社会的具体措施与实施路径。

三、研究方法

“上海推进养老服务高质量发展研究”课题组在 2023 年 10 月至 2024 年 4 月期间,前往新加坡、瑞士、日本等国家与上海、杭州等城市进行实地调研。课题组运用焦点小组、深度访谈等实证研究方法来获取一手数据资料,重点考察怎样在数字适老化改造时保留部分“老方式”,达成线上线下的有机联动,从而为上海构建老年友好型、数字包容下新旧媒介并行的微粒社会,提供一条“思维全球化、行为地方化”的全球在地化创新发展路径。

四、研究发现

通过实地走访新加坡、瑞士、日本的政府机构、社区、养老机构、社会组织、银行和医院,深入访谈当地的老年人、政府官员和社会工作者,课题组发现:在推进数字适老化改造时,这些国家有效保留了传统服务方式,深化了线上线下服务的有机联动。这些策略提高了老年人对数字技术的接受度,增强了他们的社会参与感,为上海数字技术适老化 2.0 升级行动提供了新的视角和实践范例。

（一）新加坡

1. 目标设计：构建“数字包容型社会”，而非“全民数字化社会”

在新加坡的调研发现，老年群体在数字实践方面存在显著异质性。具体来说，从完全拒绝接触数字设备的老年人，到熟练使用多种数字设备且积极推广数字产品的老年人，多种类型都有。一方面，有些老年人数字接受度较高，其数字技能和年轻群体相差无几。他们能熟练使用高频生活场景中的 APP，如手机银行、超市会员应用等，甚至用智能手表进行健康监测。另一方面，部分老年人对数字生活持观望或保守态度。一些人担心财产安全，拒绝使用任何涉及金钱交易的应用；还有些人没买手机，出门只能借用他人的设备。此外，随着老年人视力、听力和肢体协调性逐渐变弱，部分老年人已失去使用复杂数字设备的能力。

基于新加坡老年群体数字实践存在差异的特点，新加坡政府在构建数字智慧城市时，其目标更多是构建能容纳所有类型老年群体的“数字光谱”，倡导打造“数字包容型社会”，而非单纯追求“全民数字化社会”。这种数字包容思想体现在因势利异、因人而异的原则上，即尊重老年人个体差异，采用灵活且个性化的数字化转型策略，以确保每位老年人都能在数字转型中找到适合自己的位置。一方面，要保障那些完全没有能力或不愿使用数字设备的老年人的生活服务需求得到满足；另一方面，政府要提供多层次的数字教育，推动大多数老年人向数字化转型。

2. 服务设计：线上操作只需“最后一公里”，线下保留专门服务通道

新加坡在老年人生活的高频场景中，充分保留了线下服务通道，以推动数字转型的平稳过渡。本文以实地调研的银行和超市场景为例展开分析：

一是新加坡的银行在推进数字化转型时，保留了专门的线下服务通道，确保老年人能继续使用传统金融服务方式。这些线下服务通道包括以下几

方面。(1)现场服务:银行网点依旧提供传统柜台服务,老年人可在此办理各类金融业务,如存款、取款、转账等,由专业银行员工提供服务,确保老年人能得到面对面的帮助与支持。(2)优先服务柜台:部分银行网点设立优先服务柜台,专为老年人提供快速服务,减少其等待时间。这为处于数字转型中的老年人提供了更长的适应期,让他们逐渐适应新的技术平台。

二是把超市会员卡“装进”手机,老年人只需完成“最后一公里”操作即可。在新加坡,为帮助老年人更好融入数字社会,政府和企业采取了诸多措施,如在超市会员卡转至手机 APP 时提供专门协助。新加坡政府与企业合作推出“超市学习之旅”项目,派专人帮助老年人安装、使用手机 APP,简化复杂的设置步骤,让老年人只需完成“最后一公里”的操作即可。

3. 产品设计:运算从前台转入后台,实现人机无感化交互

为了让老年人在不改变生活习惯的前提下,经过简单的学习和适应过程,就能享受到智能化社会的成果,实现无感知应用,作为创新主体的企业在设计数字产品时将前台的数字交互过程转至后台,进行“科技无感化”转换。

这一思路在新加坡防控公共卫生事件时得到了成功应用。用户把基于 NFC 技术的令牌(Token)贴到 NFC 读卡器上时,读卡器会读取 Token 中的数据,再将其上传到政府服务器,服务器依据数据判断用户是否满足进入特定场所的条件。该应用借助接触式感应等物联网技术,把相关运算从前台个人操作转至后台自动化,以贴卡取代手机扫码输入信息这种烦琐的人机交互过程,避免了老年人因不会操作手机扫码输入而寸步难行的情况。

这种操作思路让“数字交互从前台转入后台,人机交互无感化”,不仅提高了老年人使用的便利性,更关键的是减轻了他们的数字焦虑,使他们能更自然地融入数字社会,在享受现代科技便利的同时不会因技术障碍而困扰。

4. 制度设计:成立新加坡数字转型办公处,全面实施“数码乐龄计划”

新加坡资讯通信媒体发展局(IMDA)在 2020 年 6 月设立了新加坡数字转型办公处(SG Digital Office,SDO)。该办公处旨在加速数字技术的普及与应用,推动新加坡构建包容性数字社会,确保人人都能成为新加坡数码社区的一员。为了助力老年人、小贩以及商家等群体更好地适应数字化生活,SDO 推出了“数码乐龄计划”,提供从基础通信技能到高级电子支付、数码银行技能的分级指导。

一是数字大使计划。招募 1000 名“数字大使”,他们由全职员工和志愿者组成,深入社区对 10 万名老年人进行一对一数字技能指导,帮助老年人学习使用二维码支付、电子政务服务等。此外,他们还活跃在 112 个市场、咖啡店和公司食堂,助力 1.8 万名摊主使用统一的电子支付方式。

二是社区援助站。在全国 30 多个地点设立数码转型社区援助站(SG Digital Community Hubs),这些地点遍布全国的民众俱乐部和公共图书馆等公共场所。援助站提供一对一指导和小组课程,以帮助年长者掌握智能手机、电子支付、政府电子服务等数码技能。社区援助站为老年群体提供了一个“数码避难所”,使他们能在无压力的学习环境中无顾虑地提问、寻求帮助,得到个性化的指导,并且根据自身学习进度获得持续支持,如课后笔记和复习材料等。

三是乐龄手机上网津贴计划。此计划与电信公司合作,推出专门为老年人设计且经济实惠的手机套餐,提供智能手机和数据流量补贴。符合条件的老年人能够以 20 新加坡元的价格购买基本款智能手机,且每月只需花费 5 新加坡元就可获取最少 5GB 的数据流量。

截至 2023 年 3 月底,已有超 21 万名年长者从“数码乐龄计划”中获益。通过这些措施,新加坡政府不但帮助老年人掌握了数字技能,还提升了他们的数字包容性与社交参与度,让他们能更好地融入数字社会。

（二）瑞士

1. 信息支持：线上有“口袋里的日内瓦”，线下有“老年城”公告栏

日内瓦是瑞士的第二大城市，其 60 岁及以上老年人口占总人口的 20%。日内瓦以“回应老年人需求，让老年人尽可能在最好的条件下于自己家中长久生活”为养老政策的总目标，在养老服务方面投入巨大，被列入世界卫生组织“老年友好城市”名录。

日内瓦政府为提升居民生活的便利性，开发了一款名为“口袋里的日内瓦”（Genève en poche）的手机应用程序。这款手机 APP 专为老年人的特殊需求设计，设有老年专属页面，提供超大字体的功能选项，以满足老年人视力方面存在的特殊需求。老年专属页面包含多个子菜单，这些子菜单与对应的政府官方网站建立了跳转链接。这样设计是为了让老年人更便捷地获取各类政务信息，如社会保障政策、医疗服务资讯等。此外，该 APP 还有一键报告街头骚扰的功能，老年人只需简单操作，就能直接向警察局登记报告相关事件，为老年人的出行安全提供了有效保障。

但与此同时，鉴于老年群体对传统纸媒仍有较强的依赖心理，日内瓦政府特意保留了社区宣传手册、“老年城”（Cité Senior）公告栏等线下信息传播渠道。日内瓦“老年城”的工作人员指出：“尽管数字化手段不断发展，可在老年群体中，传统的线下信息传播渠道依旧占据着不可替代的地位。在推广针对老年群体的信息服务时，需要综合考虑他们媒介使用习惯的多元性和复杂性。”

2. 数字支持：“一小时换平米”老少共居，对老年人进行数字反哺

瑞士日内瓦大学推出了一项名为“一小时换平米”（1h par m^2）的创新项目。该项目构建了一种基于互助与共享的居住模式，旨在强化代际交流互动、开展数字反哺并增强数字融入。在此模式下，日内瓦大学的学生可以

选择住进独居老人的住所,每周为老人提供 3—5 小时的服务,如计算机辅助教学,从而让老人掌握基本的数字设备操作技能,更好地融入现代社会生活。其他服务还包括外语会话辅导、采买购物协助、儿童保育服务等。作为回报,大学生能以低价租金获得居住权。

自 2019 年该项目启动后,每年都有六七十组“老少结对”成功。每组“对子”共同生活的时间往往长达一整年。该模式有效缓解了独居老人的“数字隔离”与“数字排斥”现象,同时解决了大学生群体的住宿难题。这一长期合作模式为促进代际互动、数字反哺,改善老年人心理健康等提供了丰富的实践样本。

3. 社会支持:线上报名线下陪读,满足知识和情感需求

瑞士的老年人大多热爱阅读,对知识和文化生活有着持续追求。日内瓦市立图书馆的馆藏布局细致考虑了不同群体,分区包括儿童、青少年、成年人、老年人、非法语母语者以及残障人士等类别。不过,它并没有刻意区分年轻人和老年人的读物,这从侧面体现了日内瓦社会所秉持的年龄“非歧视”原则。

然而,随着年龄的增长,老年人面临着视力下降等诸多生理挑战,这在一定程度上影响了他们独立阅读的能力。基于这一现实需求,陪读服务应运而生。

在日内瓦当地,“阅读与陪伴”(Lecture et Compagnie)公益组织在这一领域颇有名气。该组织由一家私立基金会运营,构建了一套较为系统的配对机制来开展陪读服务。有需求的老年人或其家属可在该组织官网填写详细的个人信息,内容不仅有姓名、年龄、出生日期、国籍等基本信息,还包含老年人会使用的语言、专业方向、职业、兴趣爱好以及不喜欢做的事等多维度信息。通过全面收集资料,最大程度地为老年人匹配到志趣相投的志愿者“读伴”。

陪读形式多种多样。志愿者与老年人共同阅读某一本书后深入聊天讨论是一种形式,志愿者朗读文章,老年人安静聆听也是一种形式。阅读的内容和地点由双方协商确定,居家伴读较为常见,这种自主性给双方带来了极大的灵活性。另外,陪读活动不只是单纯的读书,双方还能一起运动、下象棋等,重点在于提供陪伴。

就服务模式而言,这种陪读服务起初设定为一对一模式,目的是为老年人提供个性化、专属的陪伴体验。然而,在实际运行时,居家陪读服务有向团体活动方向发展的趋势。

(三) 日本

1. 自助:退而不休,在社会参与中逐步适应和接受数字技术

2013 年,日本进一步修订《老年人雇佣安定法》,强制企业废除招聘年龄限制,以保障老年人的就业机会。同时,地方政府借助“老年人才中心”,为老年人提供短期、临时性的灵活就业机会,如鼓励老年人在线销售手工艺品、提供在线咨询服务等。这些举措不仅有助于老年人延续职业生涯,还为他们提供了逐步适应和接受数字技术的平台。

笔者在日本调研时发现,老年人广泛参与各个服务业领域。在机场、海关、博物馆、超市、出租车行业、高速公路服务区等地,都能看到老年人的身影。站在街头随手拦一辆出租车,驾驶者往往是老年人,他们能熟练使用智能导航和安全驾驶辅助系统。实际上,截至 2023 年底,日本 75 岁以上的驾照持有者已达 728 万人(人民网,2024)。家住三重县度会郡大纪町大内山的藤原家庭,是老年人在数字时代自助并融入社会的典范。这个家庭由年近七旬的藤原夫妇和 64 岁的妹妹组成,他们的子女和孙辈都住得远,很少有机会团聚。为了保持经济独立并参与社会活动,藤原夫妇把房子改造成民宿,通过在线平台宣传和预订,成功吸引了游客。他们还利用智能手机和

互联网进行日常沟通、财务管理以及客户服务,不仅实现了自我管理和经济自立,还积极融入了数字社会。

这种社会参与既契合日本传统文化强调的“一生悬命”观念,也就是一个人一生都应持续努力工作,也有助于老年人更好地融入数字社会。在工作中,老年人接触并使用数字技术,逐渐适应现代工作环境,提升了自身的技术素养与社会参与感。日本的实践证明,鼓励老年人退而不休、继续参与社会劳动,不但有助于缓解劳动力短缺的状况,而且能借助实际工作中的技术应用,让老年人逐步适应并接受数字技术。这种做法凸显了数字适老化改造的社会文化维度,保证老年人在数字社会中既能得到技术上的便利,又能维持社会联系和自我价值感。

2. 公助:AI 作为“人的延伸”,公共服务沿袭传统

加拿大著名传播学者麦克卢汉曾深刻指出:“媒介是人的延伸。”(McLuhan & Gordon, 1964:17-36)这一观点在日本老年社会的公共服务实践中得到了生动的印证。AI 技术的创新应用与传统服务方式的保留相辅相成,二者共同构建了一个包容且普惠的社会服务体系,使老年人的生活更加便捷、安心,实现了“公助”目标。值得注意的是,日本在公共服务领域保留传统方式并非守旧,而是因为深刻理解并尊重老年群体的生活习惯与实际需求,力求在现代化进程中实现包容性与普惠性发展。

一是 AI 技术的创新应用。在考察的多个养老机构里,自动化和无障碍化的设施使老人们的生活更加便利。例如,特制的 24 小时监护床可通过感应分析判断老人的各项身体指标;具有互动功能的安抚物为孤独老人提供心灵慰藉;可穿戴的 AI 机器人设备能帮助行动不便的老年人或残障人士重新行走,加强康复训练。铃鹿机器人护理中心研发的智能康复机器人“外套”是目前世界上最先进的 AI 康复设备之一,它能让身体残障、基本失能的人士通过意识控制机器人做出相应动作,实现较好的康复效果。值得一

提的是,虽然 AI 技术在公共服务领域取得了显著进展,但从更广泛的范围来看,日本的 AI 技术主要应用于工业生产,如汽车、家电制造等领域。

二是保留传统服务方式。与老年群体密切相关的公共服务领域,仍然保留着大量传统服务方式。例如,在通信服务方面,公共场所随处可见传统电话亭和“黄页”;交通服务方面,纸质车票和纸质时刻表依旧常见,便于老年人查阅;医疗服务方面,诊所和医院仍然重视医生与患者面对面的直接交流,较少使用远程医疗方式,虽然电子病历系统已经推广,但纸质病历和处方仍是重要的记录手段;生活服务方面,小型杂货店和便利店遍布各地,它们不仅提供商品,还是人际交往的场所和社会支持网络的一部分;日本是全球现金使用率较高的国家之一,老年人中现金支付习惯依旧广泛存在,银行和商家也提供了相应的便利设施和服务;社交与信息服务方面,邮政系统在老年人生活中仍占据重要地位,他们常通过书信与他人联系,地方报纸和杂志也为他们提供了熟悉的信息获取途径,便利店和酒店里普遍能看到报刊;教育培训方面,老年大学和兴趣班大多采用面对面授课方式,注重实践操作和人际互动,并非完全依赖在线学习平台。

五、对上海的启示

微粒化是将大问题拆解为小场景。在数字化浪潮里,微粒化策略是解决老年群体数字融入难题的关键路径。该策略强调把宏观的数字适老化议题细化到具体生活场景中,从细节入手,保留并优化传统服务模式,推动线上线下服务无缝衔接。其目的在于构建老年友好型数字包容生态,在这个生态里,新旧媒介并行不悖,线上线下有机联动,协同推动老年人在数字时代自主参与、积极融入,达成技术进步与人文关怀的有机统一,为老年人打造既熟悉又现代的数字生活空间。

（一）培育老年友好型数字包容：涵摄技术包容和心理包容双重维度

数字包容，也称电子包容（E-inclusion）、数字融合、电子融合。2019 年，国际电信联盟（International Telecommunication Union，ITU）将“数字包容”定义为：旨在确保所有人都有平等的机会并具备相应技能，从而能够从广泛的数字技术和系统中受益的策略（杨巧云、梁诗露、杨丹，2022）。在笔者眼中，数字包容型社会是一种致力于在数字化进程中，确保所有人都能平等地获取、使用数字技术并从中受益的社会形态。“数字包容”能让我们突破阶层、群体之间根深蒂固的障碍，看到处于“掉线”状态的“数字弃民”，从而思考社会治理与公共服务的数字化转型。上海在下一步实施数字技术适老化 2.0 升级行动、建设老年友好型数字包容社会时，不仅要重视技术的普及和应用，更要强调心理层面的接纳与适应，也就是涵盖技术包容和心理包容双重维度。

1. 技术包容：尊重老年群体的主体地位，实现“科技无感化”转换

对于老年群体而言，技术包容不仅在于提供契合其生理与心理特性的技术产品和服务，更在于尊重老年群体的主体地位，构建一种全方位助力老年人融入数字社会的技术思维与技术手段。

借鉴新加坡经验，实现技术包容的关键在于尊重老年群体在数字化进程中的主体地位与参与感，推动科技无感化转换，切实营造有利于老年人智能生活的环境，破解数字服务对老年群体特有的“智能却不便利”的困局。

科技无感化转换，即借助技术手段让老年人能以自然、直观的方式使用数字服务，无须刻意学习或适应新技术。其意义在于大幅降低老年人使用数字服务的门槛，达成老年人无感知应用，使老年使用者在尽量不改变生活习惯的情况下，经过简单的学习和适应过程就能享受到智能社会的成果。

为实现“科技无感化”转换,科技企业等技术开发者可以设计并应用对老年人来说极为方便的“傻瓜式”系统(朱勤皓,2021),采用人机交互无障碍的技术手段来提高设备易用性。“傻瓜式”系统即操作简单、直观的系统,老年人无需专业知识就能轻松上手。人机交互无障碍的技术手段主要包含语音识别、自然语言处理、图像识别和手势控制等。例如,利用语音识别技术,老年人能通过语音指令操作设备,避免复杂的键盘输入;运用自然语言处理技术,可帮助老年人以自然对话的方式与智能设备交流;基于图像识别和手势控制技术,摄像头可以捕捉老年人的动作,智能设备据此自动调整显示内容或执行相应操作。

此外,应鼓励老年群体中的数字熟练使用者积极参与技术研发和服务改进过程,让数字技术更贴合老年群体的实际需求,真正达成技术包容的目标。

2. 心理包容:构建全方位立体化的支持网络,减少技术恐惧和数字焦虑

对于老年群体而言,心理包容侧重于关注老年人在心理层面对于数字技术的适应、接受与认同状况。老年群体中有大量的“数字难民”和“数字移民”,他们在技术飞速发展时会产生数字排斥、技术恐惧和数字焦虑等心理障碍,这些问题需要得到重视与解决。心理包容需要家庭、社区、政府和社会各界携手努力,营造一个容忍“慢”、允许数字“掉线”的环境,减轻老年群体对新技术的恐惧与抵触情绪。

根据社会支持理论(House, 1981),个体面临压力或挑战时,来自家庭、社区和社会的支持可显著提升其应对能力。借鉴瑞士与新加坡的经验,在老年人融入数字生活的进程中,构建全方位、立体化的社会支持体系是实现老年人心理包容的重要基础,这种多层次的支持网络有助于老年人在心理上更乐于接受和使用数字技术。

首先,在日常生活里,家庭成员应给予老年人更多鼓励与支持,让“数字家庭”成为社会支持网络的基本单元。一方面,家庭成员中的年轻人是技术社会化的推动者,可通过“代际数字反哺”,也就是年青一代向年长一代传授数字技术与知识,来帮助老年群体树立使用数字技术的信心。在上海,“代际数字反哺”模式已取得一定成效,约 65%的老年人称自己从子女或孙辈那里学到了智能手机的使用方法。随着老年群体数字素养的提高,这种反哺不仅局限于基本操作,还可涵盖在线购物、健康管理和金融理财等复杂应用。另一方面,“夫妻互学”“朋辈互学”也是值得推广的有效模式。老年夫妻可在日常互动中相互学习并分享数字技术使用经验,提升彼此的数字技能。老年朋辈之间则可通过社区活动、老年大学等形式进行学习,因为老年人在同龄人中更易找到共同话题和学习动力。

此外,社区是老年人日常生活的重要场所,应发挥积极作用。比如,组织各类数字技能培训活动,鼓励老年人相互学习、交流经验,共同跨越数字鸿沟。社区可设立专门的数字技能培训中心,定期开展面向老年人的数字素养提升活动;组织老年人数字技术学习小组并定期开展活动,分享使用心得与技巧;设置“社区数字大使”,为老年人提供一对一的数字技术辅导。

政府与社会也应给予更多支持与资源,助力老年人更好地融入数字社会。政府可以出台相关政策,给予资金支持与法律保障,推动老年数字项目的施行。例如,上海在推进“银发 e 学堂”项目时,设立专项资金,为社区培训中心提供了必备的设备与师资力量。非政府组织可通过志愿服务、公益讲座等形式,为老年人提供更多心理疏导与技术帮扶。

(二) 构建新旧媒介并行的微粒社会:从“普惠共享”到“敏捷精准”

在鼓励推广新技术、新方式的同时,保留老年人熟悉的传统服务方式,包容他们使用传统传播技术或者替代技术,着力解决好线上线下深度融合

问题。

1. 普惠共享:日常生活场景保留传统“老方式”,线上线下深度融合

老年人在新旧服务方式衔接期间面临的主要难点为:技术接受程度低、操作能力欠缺、对新技术感到焦虑以及依赖传统服务方式。研究显示,约40%的老年人在使用自助服务时存在困难(Ro & Kwon, 2024)。上海推动数字技术适老化 2.0 升级行动时,需注重在高频生活场景保留传统服务方式,以保障无法适应新技术的老年人的基本需求。例如,银行网点应继续保留人工服务窗口,尤其是针对大额转账、理财咨询等复杂业务,保证老年人能得到专业的面对面服务。同时,银行可设置专门的老年人服务区域,配备舒适座椅和大字体宣传资料,方便老年人阅读、理解业务流程。超市除自助结账设备外,也要保留人工收银台,高峰时段尤其要确保有足够工作人员为老年人提供服务。此外,超市可设置老年人优先通道,缩短老年人的等待时间。

此外,将线上与线下服务有机结合,实现无缝对接。例如,可借鉴新加坡的“超市学习之旅”项目,派遣专人协助老年人安装、使用手机 APP,让老年人在不改变生活习惯的前提下融入数字社会。继续推行“数字伙伴计划”,以“随行伙伴”“智能伙伴”和“互助伙伴”这三种方式,帮助老年人解决使用数字技术时遇到的困难。持续完善并推广“为老服务一键通”重点场景建设,如“预约就医一键成”“叫车出行一键达”“紧急救援一键通”“政策咨询一键知”等高频急难场景,使老年人仅借助传统电话机、电视机、自助服务机、便携式终端等常见设备,就能享受数字化便捷服务。

2. 敏捷精准:针对不同信息阶层的老年人,提供个性化层次化服务

对于“信息拥有者”(数字技术熟练使用者)而言,可以提供更为智能化、个性化的服务方案。例如,开发适合老年人使用的智能家居设备,赋予其语音控制和远程监控功能,方便老年人在家中进行健康管理、安全监控等

事务。此外,还可提供定制化的健康管理应用,借助大数据分析给予个性化的健康建议。例如,基于个体生理特征、生活习惯、健康状况等多维度数据,利用大数据分析和人工智能算法,构建为老年用户提供定制化饮食与运动建议的智能平台。

对于“信息未拥有者”(“数字难民”),应重点为其提供基础数字技能培训与支持。培训内容可涵盖智能手机基本操作、常用应用程序使用、网络安全知识等,培训方式可将线下集中培训与线上视频教程相结合,以确保老年人逐步掌握基本数字技能。例如,依靠老年大学、开放大学、养老服务机构等,丰富老年人数字技能培训形式与内容,如体验学习、尝试应用、经验交流、互助帮扶等。此外,还可通过“一键语音”等简化操作方式,助力数字素养较低的用户表达需求。

对于“信息中下阶层”(已触网但存在技术障碍的“数字移民”),可提供补偿性与替代性解决方案,助力其跨越技术障碍。比如,推广使用操作简便的功能手机,提供语音通话和短信服务,以保障基本通信需求。此外,在社区设置数字服务站,提供设备借用与网络使用服务,帮助老年人克服技术障碍。对于部分愿意尝试智能设备的“数字移民”,提供适配性佳的智能设备。例如,一些智能手表或手环具备健康监测、信息提醒等基础功能,操作较为简单,老年人通过语音提示和简单的按钮操作就能轻松掌握使用方法,进而避免在享受智能设备便利的同时,因操作复杂产生抵触情绪。

(三) 激发老年群体内生动力:以再就业扩大社会参与,以数字消费推动数字融入

激发老年群体的内生动力,通过再就业扩大社会参与,以数字消费推动数字融入,这是促使老年人在社会参与中深度融入数字环境、提升数字素养和生活质量的关键策略。

1. 以再就业扩大社会参与,在社会参与中融入数字环境

课题组调研发现,77. 5%的上海受访老年人有退休后再就业的意愿。在行为方面,“轻龄”老人再就业的情况格外火热,六成上海受访老年人退休后正在从事二次就业或零工经济相关工作,如咨询师、老年大学教师、瑜伽教练等。

在再就业岗位上,能够增加老年群体所处环境的媒体丰富性。老年人在沉浸式感知数字环境时,常常需要面对学习新数字技术的挑战,如手机支付、外卖订单处理等。积极适应这些变革,有助于提升他们的数字技术熟练度,拓展社交范围,增进与年青一代的互动,增强老年人的数字自我效能感与数字素养,从而使他们能够更加自信、独立地运用各类数字工具,全方位地融入数字社会。再就业为老年人提供了积极接触新兴数字技术的契机,工作与社交中的数字化学习有助于他们深度融入数字社会,保持与其他年龄层的紧密联系,构建多元包容的社会结构。再就业相关支持政策不但缓解了部分老年人的生活压力,而且为数字适老化营造了有利的环境,激励老年人积极拥抱数字生活。

为此,社会企业和产业部门应建立并完善新经济业态下多元化的劳资雇佣关系。利用弹性工作时间、共享经济平台、老人再就业平台、银发创业平台、互联网 APP 等新业态,以及时间银行等志愿服务,提升老年群体的劳动参与率和就业质量,为老年群体提供发挥余热的舞台。可借鉴新加坡、瑞士、日本三国的经验做法,鼓励超市、博物馆、高速公路服务区等服务业领域招募或返聘适龄老年人,使老年人继续拥有工作岗位,通过再就业扩大社会参与,在社会参与中融入数字环境。

2. 构建线上线下融合的消费生态,以老年数字消费推动老年数字融入

以数字消费推动老年人的数字融入,重点在于构建线上线下融合的消费生态,提高老年人的数字消费技能与社交参与度。

线上,借助微信、抖音、快手等主流平台,构建针对老年人的垂直化、社群化、视频化消费生态。构建微信群、公众号矩阵、视频号矩阵等多元渠道,利用关键意见领袖(key opinion leader, KOL)的影响力,精准吸引具有相似兴趣和特征的老年群体,实现内容与交易的无缝对接,增强用户黏性,形成完整的商业闭环。同时,老年数字消费的拓展不但要关注当下老年群体的消费需求,而且应着眼于为老龄化社会提前布局,引导老年人进行前瞻性消费,兼顾“为老”与“备老”。

线下,上海市区两级文旅局可以通过官网、微信公众号、APP 客户端等官方渠道推荐适合老年人的活动路线,社团组织负责具体实施,如 City-walk、周末郊游、人文参观、科技体验、康旅康体等活动,以此丰富老年人的生活体验。结合春节、重阳节等传统节日以及“敬老月”等特定活动,电商平台与商超可以联合推出老年专属购物节,优化界面设计,设立银发消费专区,鼓励家庭成员协同消费,实现线上下单线下体验,构建线上线下一体化的消费新模式。

这种模式有助于增强老年人的数字技能与数字自我效能感,提升他们的社交参与度和生活质量,为构建老年友好型数字包容社会提供有力支撑。

参考文献

卜卫、任娟,2020,《超越“数字鸿沟”:发展具有社会包容性的数字素养教育》,《新闻与写作》第 10 期。

方提、张潇文,2023,《“银色数字鸿沟”:城市老年人智能产品使用意向的影响因素研究》,《传媒观察》第 8 期。

国务院办公厅,2020,《关于切实解决老年人运用智能技术困难实施方案的通知》,https://www.gov.cn/zhengce/content/2020-11/24/content_5563804.htm。

库克里克,2018,《微粒社会:数字化时代的社会模式》,黄昆、夏柯译,北京:中信出版集团。

李清华,2022,《老年群体数字生活的困境、成因与对策》,《广州城市职业学院学报》第 4 期。

陆杰华、韦晓丹,2021,《老年数字鸿沟治理的分析框架、理念及其路径选择——基于数字鸿沟与知沟理论视角》,《人口研究》第 3 期。

邱林川,2019,《信息时代的世界工厂:新工人阶级的网络社会》,桂林:广西师范大学出版社。

人民网,2024,《日本再现高龄“马路杀手”多数老年人不愿上交驾照》,http://m. people. cn/n4/2024/0614/c58-21117379. html。

上海市人民政府,2021,《上海市国民经济和社会发展第十四个五年规划和二〇三五年远景目标纲要》, https://www. shanghai. gov. cn/nw12344/20210129/ced9958c16294feab926754394d9db91. html。

上海市人民政府办公厅,2021,《上海市全面推进城市数字化转型“十四五”规划》,https://www. shanghai. gov. cn/202124bgtwj/20211221/9d023714880f4ae79dc9c78fff4712d4. html。

王志伟,2021,《超七成受访者感觉父母在生活中面临诸多“数字化障碍”,74. 5%受访者会耐心为父母提供指导》,《中国青年报》12 月 9 日。

新华社,2024,《2577 家网站和 APP 完成适老化及无障碍改造》,https://www. gov. cn/lianbo/bumen/202401/content_6927154. htm。

杨巧云、梁诗露、杨丹,2022,《数字包容:发达国家的实践探索与经验借鉴》,《情报理论与实践》第 3 期。

张虎祥,2023,《以安全老龄化防范化解养老风险的探索路径——基于上海的研究》,《上海城市管理》第 6 期。

张朋辉,2021,《科技赋能,让老年人乐享数字生活》,《人民日报》1 月 4 日。

中国互联网络信息中心,2024a,《第 54 次中国互联网络发展状况统计报告》,https://www. cnnic. cn/n4/2024/0829/c88-11065. html。

中国互联网络信息中心,2024b,《第 53 次中国互联网络发展状况统计报告》,https://www. cnnic. cn/n4/2024/0322/c88-10964. html。

周煜,2022,《智能化时代美国老年数字鸿沟的现状与启示》,《国外社会科学》第 6 期。

朱勤皓,2021,《人工智能赋能下的养老服务思考》,《中国社会工作》第 23 期。

House, J. 1981, "Work Stress and Social Support." https://www.semanticscholar.org/paper/Work-stress-and-social-support-House/1d95eff51cc94bc389b4ee70c63b3ee27f1e46dc.

McLuhan, M. & W. Gordon 1964, *Understanding Media: The Extensions of Man: Critical Edition*, Berkeley: Ginko.

Prensky, M. 2012, *From Digital Natives to Digital Wisdom: Hopeful Essays for 21st Century Learning*, Thousand Oaks: Corwin Press.

Quan-Haase, A., G. Mo & B. Wellman 2017, "Connected Seniors: How Older Adults in East York Exchange Social Support Online and Offline." *Information, Communication & Society* 7.

Ro, Y. & B. Kwon 2024, "Does User Burden Matter? Age Differences in User Behavior of Self-Service Technology." DOI:10.1080/10447318.2024.2375085.

Robertson, R. 1995 "Glocalization: Time-Space and Homogeneity-Heterogeneity." *Global Modernities* 1.

人工智能中的人的形象*

——与约瑟夫·魏岑鲍姆对话

伯恩哈德·珀克森　著**

王立秋　译***

摘要:约瑟夫·魏岑鲍姆为躲避纳粹而移居美国后,致力于数学研究并成为麻省理工学院计算机科学教授。他因 ELIZA 程序而闻名。该程序的功能是模拟一位(至少在表面上)试图从心理层面理解客户的心理治疗师,成为现今模拟人类语言的聊天机器人的早期先例。魏岑鲍姆的研究使他逐渐对计算机的可能性、局限性和用途持批判态度。他关于计算机对人类经验世界影响的开创性著作《计算机的力量和人的理性》,触及了一个在当时全新、很大程度上未知且即将迎来爆发式增长的领域。本文为魏岑鲍姆在 1998 年接受的一次访谈,主要关注人工智能的发展、人与人工智能之间的类比论证,以及人机关系的批判性思考。

关键词:人工智能　约瑟夫·魏岑鲍姆　ELIZA 程序　人机关系　技术伦理

一、约瑟夫·魏岑鲍姆:生平概述

魏岑鲍姆生于 1923 年,父母都是犹太人,1936 年离开纳粹德国移居美

---

* 本文原题为“The Image of Man in Artificial Intelligence: A Conversation with Joseph Weizenbaum”原刊于 *Weizenbaum Journal of the Digital Society* 3(3),由作者以 CC-BY 协议许可翻译发表。

** 伯恩哈德·珀克森(B. Pörksen),德国图宾根大学媒体研究教授。

*** 王立秋,哈尔滨工程大学人文社会科学学院。

国,后致力于研究数学。在底特律的韦恩大学担任研究助理期间,他开始与计算机合作,编写 Slip 编程语言,并于 1955 年为美国银行设计计算机系统,随后在斯坦福大学担任客座讲师。1963 年,魏岑鲍姆成为麻省理工学院计算机科学教授,并任教至 1988 年退休。他因 ELIZA 程序而闻名。该程序的功能是模拟一位(至少在表面上)试图从心理层面理解客户的心理治疗师,其提问和回答方式类似于卡尔·罗杰斯学派的心理治疗师。

魏岑鲍姆在越战期间积极参与和平运动,并在此后持续参与此类活动。他是德意志联邦共和国计算机专业人员社会责任组织(Computer Professionals for Social Responsibility)和计算机科学家和平与社会责任论坛(the Forum of Computer Scientists for Peace and Social Responsibility)的联合创始人。1976 年,他出版了开创性著作《计算机的力量和人的理性》(*Computers Power and Human Reason*),论述了计算机对人类经验世界的影响。在当时,这是一个全新且迅速发展的领域,很大程度上尚未被探索。在另外两本书《驶向冰山》(*Kurs auf den Eisberg*)和《谁发明了计算机神话》(*Wer findet die Computermythen*)中,他将注意力转向军事与计算机科学研究之间的联系,并对人工智能(AI)的支持者提出了强烈批评。

二、个人的回忆:作为科学的批评者、神秘主义者和人的约瑟夫·魏岑鲍姆

当我在这篇访谈发表之际回想约瑟夫·魏岑鲍姆时,与他的交谈让我感到无比兴奋和满足。[①] 有三次与他接触的经历让我记忆深刻。

首先交代背景。在我年轻时,魏岑鲍姆是我父母在弗莱堡家中的常客。

① 原访谈最初以德文发表于 2000 年的 *Communicatio Socialis* 1, pp. 4–17,并收入 Flessner, 2000。顺便说一句,早在 1997 年,我就已发表过一篇对魏岑鲍姆的访谈,参看 Pörksen, 1997。

他常在晚餐时兴致勃勃地嘲弄科学界,有一次还在饭桌上批判了受路德维克·弗莱克、托马斯·库恩发展的知识社会学影响的范式力量(Fleck & Ludwik, 1993; Kuhn, 1991)。在一个特别难忘的场合,他即兴编造了一个认识论寓言,这一行为反映了他的个人特质:没有恶意,但思维敏捷、爱开玩笑,且乐于讲故事。他将这个寓言称为"一个反例的简历"。听他讲话时,我意识到这位科学家的思维方式与众不同。尽管不久前他还被公认为天赋异禀、富有创新精神,但他却提出了一个与主流科学和常规科学所接受的真理相悖的反例。对魏岑鲍姆来说,我们可以对这个反例提出几个问题。首先,这个如此大胆的人是谁?即使是一个简单的问题,也足以削弱反例的可信度。毕竟,谁会相信一个无名小卒,花时间听他说话呢?其次,同样重要的是,这个相信自己能动摇既有范式的人来自哪里,或者说来自哪个思想流派?最后,这个人在哪里发表文章?在哪些期刊上,与谁合作?是受认可的作者或知名期刊吗?也就是说,我们真的需要认真对待这个人吗?魏岑鲍姆真正想表达的是,科学知识的生成过程从来不只是理性的,它总是掺杂着社会因素。问题在于个体、流派和团体对权力的影响,以及一种难以捉摸的、分散的声誉评估(这种评估最终以规训、要求服从,有时甚至排除某些人为结果)。我意识到,约瑟夫·魏岑鲍姆正在以自己的方式接受这些经历:作为一名讲故事的学者,含蓄而机智、不死板,且充满幽默。

让我记忆犹新的第二段经历发生在我的学生时代,而且充满了令人惊奇的荒谬感。哲学和美学教授鲁道夫·利珀亲王(Prince R. zur Lippe)邀请魏岑鲍姆到风景如画的于德庄园参加一个为期数天的研讨会。由于邀请匆忙,且分组讨论安排在周中,大部分受邀者未能出席。这让我和一位大学朋友有幸进入邀请名单,并在庄园的冬园里与魏岑鲍姆聊了几个小时。其间,孔雀和天鹅在窗前漫步,偶尔好奇地看着我们。对我们来说,这是一次难得的机会,令人激动不已。魏岑鲍姆看到参会人数稀少,起初有些不悦。

第一轮讨论后,他调侃利珀,下次可以在“电话亭”见面,那里的空间肯定足够。傍晚,我们在利珀的餐桌上用豪华餐具享用了附近餐馆送来的烤兔肉。几位与会者在餐中发现了子弹,小心翼翼地将它们放在盘子边缘。尽管如此,气氛依然轻松。魏岑鲍姆在餐后讲话中突然提到神秘思维对他工作的重要性。尽管这座曾是修道院的庄园显得庄严,他却以迂回、间接的方式谈论神秘,仿佛是否定神学的代表。他用谜语般的方式围绕这一主题,但坚定地表示,“有些东西不可说,活的真理无法用语言表达”,正如他在这篇谈话中提到的那样。那天晚上,与他在麻省理工学院和自然科学研究中的思维形式不同,他以一种令人印象深刻的、对理性的审慎态度(而非流行的对理性的敌意)强调,世界从根本上说是无法被完全解释的。那一刻,我清晰地感受到,他不仅仅是一名计算机科学家,更是一名试图在一个被认为可以被看透的世界中,为惊奇感找到立足之地和存在理由的神秘主义者。

我想回忆的第三段经历发生在不来梅。文化批评家、技术哲学家、《欢乐相会的工具》(*Tools for Conviviality*)的作者、曾担任教宗蒙席的伊万·伊里奇(I. Illich),与他的伴侣、身体历史学家芭芭拉·杜登(B. Duden),定期在家中举办聚会。周五下午,他在不来梅大学授课后,常邀请学生和来自世界各地的朋友到家中继续讨论至深夜。在这里,我也多次与魏岑鲍姆见面。1998 年,伊里奇获得不来梅伊洪别墅文化与和平奖(Culture and Peace Prize of Villa Ichon in Bremen)。受病痛折磨的伊里奇在获奖致辞中,以深刻而忧郁的方式描述了一个知识分子和世界公民对“家”的追寻。[①] 对他而言,在经历了无数次停留——包括在拉美和美国的长期居留,以及在世界多所大学担任客座教授——后,最终在不来梅找到了一个“家的角落”[②]是一

① 伊里奇同样为躲避纳粹而逃离,他的母亲在皈依天主教前是犹太人。他逃往佛罗伦萨和罗马,后在纽约成为牧师,并在 30 岁时担任波多黎各天主教大学副校长。

② 对此用语的分析,参看 Hartch, 2015。

种意外的幸运。那一刻,他像诗人和神秘主义者一样说话,仿佛成为另一位否定神学家。活动结束后,我与约瑟夫·魏岑鲍姆站了一会儿,从未见过他如此激动。他只说了一句:“那太棒了!”我意识到,他也是一位流离失所者、一位寻找者,最重要的是,一位回到德国并在出生地柏林度过余生的调解者。以下谈话发生在 1998 年那次相遇后的某个时候。我希望它能向读者展示作为科学批评者和神秘主义者的约瑟夫·魏岑鲍姆特有的人性。

三、人工智能中的人的形象:与约瑟夫·魏岑鲍姆对话

(一) 机器的模型和人的模型

珀克森:任何与人工智能打交道的人,都会反复遇到两种截然不同的视角:一些人工智能支持者将人视为机器的模型,认为机器智能应模仿人类智能;然而,另一群研究者则认为,人类已不再是理想的模仿模型,未来的机器将超越人类,成为近乎完美、拥有神奇思维力量的实体。您能确认这两种观点中哪一种是最初就存在的吗?

魏岑鲍姆:自 1956 年达特茅斯会议提出“人工智能”概念以来,这两种观点就以不同形式并存。从一开始,不仅人类被视为模型和理想,机器也被认为是待建立的模型模式。一方面,马文·明斯基(M. Minsky)曾说:“大脑只是一台肉做的机器。”在英语中,德语 Fleisch 有两层含义:肉(meat)和肉体(flesh)。肉体是活的,而肉则是死的。你可以随意处理肉——吃它、煎它、烧它。然而,明斯基的这句话依然以人类为榜样,人们希望复制人类的智能。这一观点隐含了一个想法:复制人类智能并不难,因为人类的核心器官——大脑——不过是一台肉做的机器。

另一方面,明斯基还提出另一种观点:神作为工程师并不完美,人类是一种存在许多多余缺陷的异常现象。人类需要睡眠,花费大量时间学习却会遗忘,最终还会死亡。然而,根据早期在人工智能社区中流行的看法,我们可以创造出某种更好、更完美的东西。我们正在构建的理想模型在某些方面仍以人类为导向,但它更接近完美。

珀克森:进一步推演这一想法,它实际上意味着,从这一角度来看,人类将成为连接猿人与超越人类的机器存在之间的进化环节。布鲁塞尔自由大学的人工智能教授吕克·斯提尔斯(L. Steels)建议将这种未来存在称为"赛博智人"(homo cyber sapiens)或"人智机器人"(robot hominidus intelligens)。

魏岑鲍姆:卡内基梅隆大学移动机器人实验室主任汉斯·莫拉维克(H. Moravec)在《心智后裔》(*Mind Children*)中提出了类似观点。他相信,人类意识可以下载到机器人身上,我们可以而且应当复制人类意识,使其脱离必死的身体。在他看来,身体只是心智的载体,而人的本质是信息。他对身体持鄙视态度。

珀克森:其他与莫拉维克观点相近的作者也毫不客气地将身体称为"一块肉""软乎乎的东西"或"罐头肉"。

魏岑鲍姆:这只是这个圈子的表达方式。无论如何,从这一角度看,未来的机器人代表对自然的修正和改进。莫拉维克声称,这样的机器可以"推动我们的文明进化"。他希望有意识地、控制性地干预他认为有缺陷的进化过程。值得一提的是,在美国,他的想法并未被视为科幻,也未受到忽

视。这些观点在顶尖大学中广泛传播——毕竟,他的书是由哈佛大学出版社出版的。

(二) 普罗米修斯的羞耻和普罗米修斯的骄傲

珀克森:我们可以认为,这种对作为“创造之冠”的机器的尊崇表达了一种基本情感。哲学家君特·安德斯(G. Anders)将这种感觉称为“普罗米修斯的羞耻”,即一种在自我设计的机器面前感到低人一等、无助甚至羞耻的心理。您同意吗?

魏岑鲍姆:这种普遍的心理现象——在面对特别完美的性能时,人们会感到低人一等、无助甚至羞耻——确实存在。从这个意义上说,君特·安德斯是对的。

珀克森:马文·明斯基曾说,硅脑的思维力如此神奇,以至于即使未来的机器人将我们当作宠物饲养,人类也会感到幸福。

魏岑鲍姆:但这不仅仅是羞耻的问题。我们可以观察到一种程序员熟悉的辩证情感:即使你认为计算机更优秀、更聪明,但你仍会感到某种骄傲。毕竟,你是编写那个特殊程序的人。于是,一个基本矛盾出现了:你为自己的弱小感到羞耻,同时又为能够创造出如此高级的程序而自豪。

珀克森:莫拉维克称新机器人是他的“精神后裔”,这种说法也与这种骄傲感相关。

魏岑鲍姆:令人惊讶的是,一个人竟会以这种方式说话。这是巧合吗?

我不这么认为。毕竟,人工智能领域是由男性主导的。在我看来,除了扮演神的妄想外,对女性及其生育能力的嫉妒也是一个驱动因素。我称这种现象为“子宫羡妒”。

珀克森:您是说,这是西格蒙特·弗洛伊德描述的“阴茎羡妒”的补充现象?

魏岑鲍姆:没错。现在你可以假装自己也能生育,甚至比任何人类都做得更好、更聪明。

珀克森:但这种说法完全是臆测。从目前生产的机器人来看,明斯基的说法——即使未来的机器人将我们当作宠物饲养,人类也会感到幸福——似乎也过于夸张。人工智能研究相对缺乏成功案例本身不就意味着我们应该形成一种新的自我意识,即人类能看到机器无法做到的事情?

魏岑鲍姆:当然,看到那些未实现的人工智能预言,你可能会说:“难道我们不奇妙吗?难道我们不伟大吗?”但这些话语都是无意义的。我们不应以机器的能力或计算机下棋的水平来衡量人类的尊严。这种思维形式——尽管仍带有人本主义色彩——的致命缺陷在于,机器成了衡量人类的标准。机器人的能力或局限性,都不应成为定义人类自我意识的基础。

(三) 技术世界的秘密人类学

珀克森:对我来说,是否应严肃对待人工智能寓言领域的各种预测仍是一个问题。虽然大多数预测看似荒谬,但有些却有趣、令人兴奋。你会遇到各种疯狂的推测、离谱的观点和令人震惊的想法。那么,您认为有必要严肃

对待明斯基或莫拉维克的想法吗?

魏岑鲍姆:这实际上是两个问题。第一个问题是,我们是否应预测人类有能力制造出这种更高级的机器。我们应该为此感到害怕吗,即某天这样的机器人真的出现了,而我们将成为它们的宠物?我的回答是,不必害怕。此外,与这些人争论技术可行性几乎没有意义,那是一场必输的游戏,只会让这些想法朝新的、可能同样不可取的方向发展。第二个问题是,我们是否应严肃对待这一领域传播的思想。对此,我的回答是肯定的!因为与其他学者一样,这些学者在新千年来临之际宣扬了一种极端危险的人类形象。

珀克森:这种人类形象是什么样子的?

魏岑鲍姆:它基于这样一种观点,即人类是可以被理解、破译、修正和改进的机器——至少在原则上,在不远的将来可以。这种观点的核心教条是,生命的所有方面都是可计算的,可以被分解为可计算的、可公式化的进程。

珀克森:如果我没理解错,您主要关注的不是技术的可行性,而是伴随技术乌托邦出现的隐性意识形态。马歇尔·麦克卢汉所说的"技术的符号效应"与此类似。他认为,技术创造了一个环境和符号世界,从而改变了人类的自我理解。

魏岑鲍姆:即使未来有可能制造出这样的智能机器,那也将是遥远的未来、几代人之后的事。然而,今天,人类形象正在发生变化,并取得了巨大成功。明斯基和莫拉维克并非凭空出现,他们是在一个绝对相信自然科学和现代技术的文化与时代中提出这些观点的。

珀克森:我再问一次:基于机器类比的人类形象的危险性在哪里?

魏岑鲍姆:从本世纪(20 世纪)——也许是最野蛮的世纪——我们可以看出,这种人类形象在过去的罪行中起到了决定性作用。想想那些最残暴的罪行,是如何通过施害者否认受害者的人性而成为可能的。在纳粹时代,犹太人被描绘为害虫,这一比喻使大规模屠杀合法化。今天,在自然科学权威的支持下,这种观点——人类不过是可被机器人取代的信息处理机器——正变得越来越牢固。这一视角与莫拉维克关于"后生物社会"中机器人统治的论述是一致的。

(四) 论一个比喻带来的后果

珀克森:您的意思是,动物的比喻与机器的比喻相似?

魏岑鲍姆:没错,尤其是在一个关键方面相似。这些比喻摧毁了对人的尊重,使人能够接受终结生命的可能性。就像我们不会尊重害虫一样,我们也不必尊重"肉做的机器"。这样的比喻是犯下一切可想象罪行的起点。在我们的世纪(20 世纪),这一点已以可怕的方式得到证实。

珀克森:您对莫拉维克的《心智后裔》做出了特别愤怒的反应,称他的终极目标是实现"人类问题的最终解决方案"。为什么要提到纳粹的罪行?我不太理解这种类比的目的。

魏岑鲍姆:或许有些过激了。但我们都是历史的产物,历史在我们出生前就已开始。我们的经验塑造了我们,出生时的社会情境也影响了我们。

因此,作为“德国犹太人”,我被这一类比吸引并不奇怪。其他人可能会注意到不同的东西。

珀克森:但问题依然存在:纳粹的种族狂热与人工智能教授的奇想有何共同之处?

魏岑鲍姆:当然,汉斯·莫拉维克不会每天早晨对着镜子行纳粹礼。这不是重点。我想说的是,非人的人类形象具有巨大力量,这种形象会像病毒一样在社会中传播。我相信,纳粹与莫拉维克的想法本质上的共同点在于对人的贬低,以及不惜一切代价创造完美新人的奇想。然而,在这种完美的尽头,人类消失了。在莫拉维克的后生物社会中,人类不复存在。我深信,今天我们必须捍卫人道的人类形象,使其免受机器比喻的支配。

珀克森:但没有人会无助地任由机器比喻决定的人类形象所摆布。毕竟,人类的感觉截然不同,人是作为个体行动和接受的。

魏岑鲍姆:我的看法不同。个体非常脆弱,纳粹时代彻底屈服者的形象很容易证明这一点。只需回想德国人的叫喊、欢呼和“胜利万岁”的口号就够了。人们确实——尤其是在本世纪(20 世纪)——充满激情地一再屈服,直到为时已晚,最终只能惊恐地站在万人坑前。

(五) 人工智能的信条

珀克森:来自机器世界的比喻无处不在且一直存在。它们甚至可以追溯到蒸汽机、齿轮和螺丝的时代。我们会说心脏像泵,身体像钟表一样运作,或者说人的发条松了、螺丝没上紧。电力时代又引入了“能量不足”或

“保险丝烧断了”的概念。

魏岑鲍姆:事实上,这种关于世界和人的机械论观点已有大约 300 年的历史。在声称这些比喻一直存在时,我会非常谨慎。几个世纪前,但丁的地狱形象和人们凭借努力或神的善意(死后)进入天堂的观念还被当作现实。不过,我确实认为比喻无处不在。我认为,一切知识和理解事物的尝试本质上都是比喻。我们通过类比、比较和比喻来认识世界。我们在试图理解某事物时,总是使用类比模式,通过建构相似点,在已知事物的框架内把握陌生或未知的东西。

珀克森:那么,我们是否可以将马文·明斯基关于“肉做的机器”的比喻理解为一种理解大脑的尝试?

魏岑鲍姆:当然可以。但如果我说“大脑不过是肉做的机器”,情况就不同了,这一点至关重要。你会认为这是一个完整、全面且充分的描述。我的批评与使用特定类比或比喻来描述大脑或人无关。事实上,这是一个常见的过程,至少自牛顿以来的科学研究都有这种做法。但我反对的是这种论断:大脑不过是一台肉做的机器,人不过是一个信息处理系统。

珀克森:您反对的是简化论?

魏岑鲍姆:我对人工智能研究的这一信条感到被冒犯和恶心,即人类生活的方方面面都是可计算、可破解的。当然,人类会处理信息,这些比喻作为科学工作和知识的简化模型具有重要价值。它们是必要的,有助于理解人类生活的某些方面。然而,它们是对现实的抽象,永远无法完全把握整体。

珀克森:您能想到其他语言或描述形式吗?

魏岑鲍姆:如果我有这种权威,我会建议从所有科学文本中删除“不过是……而已”的表述。重要的是,现象、事物、感觉和行动永远是多样的,可以用多种截然不同的方式描述。只有在特定语境中,它们的意义才会显现。举个简单的例子:感到一只手搭在肩膀上是什么意思?只有在讲述整个故事时,才能回答这个问题。假设一个年轻人与女友吵架后,正伤心地坐在图书馆试图工作,突然感到一只手搭在肩上。或者,假设这个年轻人是被通缉的罪犯,坐在火车站候车室时感到一只手搭在肩上。我想说明的是,没有什么可以被一劳永逸地描述。一切描述都依赖于特定语境。

(六)从魔像到机器人

珀克森:在我们关于必要与危险的比喻和习语的争论之后,我想尝试以另一种方式为马文·明斯基和汉斯·莫拉维克辩护。毕竟,人造人的梦想由来已久。根据我的研究,莫拉维克和明斯基只是用现代手段实现了布拉格的勒夫拉比(R. Löw)制造魔像的目标——如果你相信传说的话。他们追随了帕拉塞尔苏斯(Paracelsus)创造人造人荷蒙库鲁斯(Homunculus)的奇想,并处于可追溯至启蒙时代的自动机狂热(automaton mania)传统中。

魏岑鲍姆:你的问题隐含了“不过是……而已”的简化论表述。你实际上是在说,莫拉维克和明斯基的梦想不过是人类历史上一直存在的梦想而已。然而,皮格马利翁的理念虽然古老,甚至可以追溯到人类起源,但这并不能证明莫拉维克和明斯基的梦想无害。相反,它恰恰证明了这种梦想的力量。它表明,理念——即使是错误的理念——具有巨大的影响力。与过

去不同的是,如今人们相信,现代自然科学的信仰将使我们真正理解人类,而高性能的计算机似乎是实现这一目标的工具,最终使今天人们欣快且傲慢地倡导的皮格马利翁之梦成为现实。

珀克森:人工智能的拥护者真的如此傲慢和欣快吗?联结主义(connectionism)这一不那么简化论的思想流派正在逐渐扩大其影响力。它将智能视为一种涌现现象,试图通过构建能够学习的神经网络来实现智能的涌现。从这个角度看,大脑是一个基于无数连接的巨大神经网络。因此,受早期控制论启发、按自组织模型训练的机器智能,也可能具有某种神秘且不可分解的复杂性。

魏岑鲍姆:我不认为联结主义者会表现出新的谦逊。马文·明斯基也曾改变观点,承认人工智能领域早期的一些方法根本是错误的,尽管它们可能带来巨大成果。必须承认,理解语言时语境至关重要;也必须承认,人类身体对我们的智能至关重要。最终,这一点会变得明显:设计让机器"看起来智能"的程序几乎没有意义。然而,承认错误并不会带来新的态度或谦逊。相反,认错被当作巨大进步来宣扬。同时,人们仍暗中希望有一天能理解人类存在,并构建出与大脑对应的神经网络——下一步则是破解它。

珀克森:在您对人工智能的批评中,值得注意的是,您实际上认为生命和人类存在是某种根本上神秘的东西,您希望这种神秘性得到承认。

魏岑鲍姆:世界充满神秘,而人工智能界的信条——"一切皆可计算"——否定了生命的神秘性。它创造了一种完全透明的错觉,暗示我们存在的所有方面都可以被拆解和解释。从这一视角看,相信奇迹和神秘似

乎只是一种愚蠢。这种“完全可预测”的断言深深伤害了我。

（七）神秘与秘密

珀克森：在您的著作《计算机的力量和人的理性》接近结尾处，有一段关于与奇迹和神秘相遇的文字：“当我们的孩子还小时，我有时会和妻子在他们睡觉时站在床边看着他们。我们心照不宣地无声交流；这一场景在人类历史上重复了无数次，几乎与人类本身一样古老。正如尤内斯库（Ionesco）在日记中所说的：人可以通过语言表达许多东西，但语言无法传达活的真理。”

魏岑鲍姆：的确，我的观点是，有些东西不可言说，活的真理无法用语言表达。

珀克森：神秘主义者也会以同样的方式谈论神。

魏岑鲍姆：我想讲一个小故事。曾有一位圣公会牧师在麻省理工学院工作，他的名字很有趣——斯科特·帕拉迪斯［帕拉迪斯（Paradise，意为“天堂”）］。我们相识多年并成为朋友。有一天，我举办了一个小型派对，斯科特也是客人之一。我女儿与他交谈后，惊讶地跑来告诉我：“斯科特说你是一个特别虔诚的人，一个神秘主义者。”

珀克森：那么，您在反对人工智能研究和日常生活计算机化时的犀利态度，是否有宗教基础？

魏岑鲍姆：哎，我并不想为犹太-基督教信仰体系或有组织的宗教辩

护。我不认为天上有个裹着床单的老人,被长翅膀的金发女人环绕,注视着世间的一切。但我的确在生命中体验过奇迹:悲伤与震惊、清晨的突然喜悦,以及人与人之间的爱。奇迹就是奇迹,无法用语言描述。要探讨它,你需要成为艺术家。

珀克森:对您来说,神不是人。那么他是什么?

魏岑鲍姆:我的回答很简单:神就是爱。我用"爱"这个字来翻译他的全能、恩典和其他属性。按此理解,神无处不在的想法就变得合理了。对我来说,这意味着神也在奥斯威辛,因为在那里,也有人出于爱试图帮助他人,哪怕只是将手搭在别人肩上安慰他们。

珀克森:你是否有任何亲身体验想分享?

魏岑鲍姆:在 20 世纪 30 年代的柏林,当犹太儿童被迫辍学离开普通高中时,12 岁的我被送到一所犹太男校。在那里,我第一次听说意第绪语,并遇到了东欧犹太人。他们住在"仓房区"(Scheunenviertel),生活极度贫困。当时,我爱上了一个犹太男孩,他衣衫褴褛,我和母亲尽力帮助他,时常带些东西给他。这次相遇的氛围让我印象深刻,我觉得用"爱"这个字来描述它最为恰当。

珀克森:后生物时代的策划者很可能会嘲笑这种对不可言说之物、对爱的神秘的尊重。例如,加州作家、机器人时代策划者马克斯·莫尔(M. Moore)将宗教描述为"一股与我们后人类社会对立的势力"。

魏岑鲍姆：在某些圈子里，人们会自我欺骗，试图忽悠他人，或只是胡言乱语。许多现代人认为，相较于宗教观念，自然科学为知识提供了更坚实的基础。对此，我的看法截然不同。在我看来，自然科学不过是今天的主流世界宗教，它有自己的新信徒（学生）、教会和教堂（大学），以及牧师和异端。它甚至有自己的“红衣主教”（诺贝尔奖得主）和具体的仪式，用以证明一个人的虔诚和忠诚。还有那些不顾日常体验、毫无保留地信仰自然科学的所谓“平信徒”。例如，大多数人相信地球绕太阳转，而非相反。但他们无法证明这一点，尽管这与他们的日常体验相悖——在日常生活中，他们看到的是静止的地球和移动的太阳。然而，他们依然盲目且不加限制地相信自然科学。支撑这种信念的——在人工智能和遗传学的帮助下，自然科学终将创造出一种人造的、更优秀的人类——正是这种类似于宗教的狂热。随着技术的每一次新胜利，这种信念都会继续增长。

（八）生者不可避免的个体性

珀克森：技术的胜利和人类变革的可能性，迫使我们重新回答哲学人类学的基本问题：什么是人？

魏岑鲍姆：我建议换个方式提问：什么是人的本质？人与可制造的机器之间的根本差异是什么？人类与其他存在有何不同？虽然听起来可能有些傲慢，但答案其实很简单。我一再强调，每个人都是其个人历史的产物。他们有独特的生平经历，生活在特定的时间点，并不可避免地嵌入周围的社会。也就是说，每个人都是独特的个例。即使未来出现与人类行为相似、外观惊人的智能机器——这一想法本身就很疯狂——它也永远不会成为人。因为它的历史不同，它必然与人不同。

珀克森:目前,我们可以预测有机与无机成分的混合。通过使用心脏起搏器、假体、助听器、人造髋关节和心脏瓣膜,我们已经将电子设备植入身体。科学史家和女性主义学者唐娜·哈拉维(Haraway, 2010)在其著名的《赛博格宣言》中庆祝了这种人与机器的融合,并敦促我们享受其解放力量。她的论点受女性主义启发,尖锐地指出,人与机器的融合、赛博格的创造,消除了男女差异的基础,即压迫的根源。如果我们都成为赛博格,继续将女性视为他者或被压迫对象就站不住脚了。哈拉维写道:"赛博格政治建立在噪声和废弃之上,它赞扬动物与人的非法融合。这种联结使男女变得复杂,破坏了欲望结构、语言和性别生产的想象权力,从而颠覆了西方认同的再生产模式,即自然与文化、镜子与眼睛、仆人与主人、身体与心灵的对立。"

魏岑鲍姆:这是一种漂亮但肤浅的想法。我注意到,处理此类问题的深刻思想家往往极其天真。唐娜·哈拉维的想法让我困扰的地方在于:首先,我不认为男女差异已从根本上得到了解决。在大多数科幻作品中,赛博格也有性别,因此在论证时不应忽视这一点。其次,你引用的那段文字展示了另一种对进步的欣快,即希望通过机器解决社会问题,将男女之间的永恒冲突转移到技术领域,并声称有技术解决方案。虽然我能解数学方程,但人与社会的问题永远无法通过这种方式得到彻底解决。

珀克森:但从其他角度看,我们是否也应感谢机器?一家电视台曾请作家彼得·格拉泽(P. Glaser)用十句话对人类发表演讲。他的最后一句话让我印象深刻:"对机器好一点!"对我来说,这意味着,尽管机器无法解决两性战争,但它们仍是知识的助手,促使我们进行人类学反思,并让我们以感激和友好的态度对待它们。

魏岑鲍姆:我也喜欢这句话,但我的理解略有不同。如果我们不鄙视机器,而是善待它们,也许我们也会以不同的态度对待人类同胞——不一定如此,但至少存在这种可能。这句话意味着,我们首先应保持友好——无论是面对机器还是他人。这位作家说:"对机器好一点!"你也可以说:"对世界好一点!"

参考文献

Fleck, L. & T. Ludwik 1993, *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv. Mit einer Einleitung herausgegeben von Lothar Schäfer und Thomas Schnelle*, [S. l.]: Suhrkamp.

Flessner, B. 2000, *Nach dem Menschen: Der Mythos einer zweiten Schöpfung und das Entstehen einer posthumanen Kultur*, [S. l.]: [s. n.].

Haraway, D. 2010, *A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s*, Cambridge: Cambridge University Press.

Hartch, T. 2015, *The Prophet of Cuernavaca: Ivan Illich and the Crisis of the West*, Oxford: Oxford University Press.

Kuhn, T. 1991, *Die Struktur wissenschaftlicher Revolutionen*, [S. l.]: Suhrkamp.

Pörksen, B. 1997, "'Die Verletzbarkeit des Individuums ist groß': Ein Gespräch mit Joseph Weizenbaum." *Universitas: Zeitschrift für interdisziplinäre Wissenschaft* 1.

计算的限度*

——约瑟夫·魏岑鲍姆与聊天机器人 ELIZA

大卫·贝里　著**

王立秋　译***

摘要:约瑟夫·魏岑鲍姆于 20 世纪 60 年代开发的 ELIZA,堪称有史以来最具影响力的计算机程序之一。ELIZA,尤其是其最著名的人格 DOCTOR,至今仍是程序员的灵感源泉。它引发了关于人工智能(AI)更广泛的讨论,催生了诸多对其的模仿。作为所有对话界面和聊天机器人的鼻祖,ELIZA 尤其令研究 AI 和计算领域的工程师、历史学家和哲学家着迷。ELIZA 能凭借相对少量的计算机代码生成类人回应,这一能力为大量类似程序铺平了道路。这些程序以对话智能体和其他人机界面的形式出现,在计算机科学领域催生出数个全新的研究领域。本文将探讨魏岑鲍姆对 AI 的贡献,并结合生成式 AI(如 ChatGPT)的当代发展状况,思考其更具批判性的著作。此外,本文还将通过研究有关 ELIZA 的讨论,为当前围绕机器学习和 AI 的争论提供一些思路。

关键词:魏岑鲍姆　ELIZA　人工智能　机器学习　聊天机器人　ChatGPT

一、引言

人工智能(AI)承诺采集并重塑人类最强大的能力——语言、创造力、

---

* 本文原题为“The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot”,原刊于 *Weizenbaum Journal of the Digital Society* 3(3),由作者以 CC-BY 协议许可翻译发表。

** 大卫·贝里(D. Berry),萨塞克斯大学媒体、艺术与人文学院。

*** 王立秋,哈尔滨工程大学人文社会科学学院。

推理和智力的本质。进步正飞速发生,在过去短短十年间,AI 就产生了实际影响,自 2019 年以来,这种发展更是显著加速。2022 年,像 OpenAI 的 ChatGPT 和 Stability AI 的 Stable Diffusion 这样的突破,被认为以前所未有的方式提升了人类在语言和艺术方面的创造力与生产力。在本文中,我想联系 ELIZA 来思考这一领域近来的进展。在许多人看来,ELIZA 是如今所谓生成式 AI 的前身,它是一个相对简单的聊天机器人(对话智能体)程序,使计算机内基于对话的界面成为可能。

ELIZA 由约瑟夫·魏岑鲍姆于 20 世纪 60 年代开发。可以说,它是有史以来最具影响力的计算机程序之一。ELIZA,尤其是其最著名的人格 DOCTOR,仍然吸引着程序员,并催生了各种对它的模仿。虽然 ELIZA 影响了计算机科学乃至更广泛的文化,但它的初始源代码从未被公布或广泛传播。[①] 不过,它是所有聊天机器人的前身、对话式人机互动的鼻祖,也是大众对计算机可能形象的想象(如 1968 年电影《2001:太空漫游》中的 HAL9000 和 2013 年电影《她》中的操作系统)的灵感来源。这个所有对话界面和聊天机器人的始祖仍然令研究人工智能和计算的工程师、历史学家和哲学家尤为着迷。值得注意的是,ELIZA 在对计算历史产生实质性影响的同时,也使其永远陷入与性别和阶级的问题性假设的纠葛之中(见下文)。ELIZA 能够利用相对少量的计算机代码生成类人回应,凭借这一能力为大量类似程序铺平了道路。这些程序以对话智能体和其他人机界面的形式出现,在计算机科学内部激发了几个全新的研究领域(Boden, 1977)。

值得注意的是,OpenAI 在 2022 年发布 ChatGPT,为复杂大语言模型(LLM)系统 GPT-3 设计合适界面时,决定采用聊天机器人的形式。虽然

① 作为一个 2021 年形成的国际研究团队的一部分,我们已经在 MIT 档案馆中重新发现了 ELIZA 丢失的用 MAD/SLIP 语言编写的代码。关于这一进展的细节已经发布于 ELIZA 档案网站,参见 https://sites.google.com/view/elizagen-org/the-original-eliza。

该公司尚未充分解释使用这种基于文本的模型背后的逻辑,但使用该系统本身就表明,很可能是创造人格的能力促使开发者这样设计系统。为进一步探究这一点,本文认为,回顾过去对理解像 ChatGPT 和类似产品这样的 AI 产品的特殊性或许有用。这就意味着要思考以前的系统,尤其是 ELIZA。我尤其想要考察约瑟夫·魏岑鲍姆(Weizenbaum, 1976)在《计算机的力量和人的理性》中呈现的反思。在书中,他联系人类思想不可简化(不能把人的思想简化为逻辑函数,尤其是简化为 ELIZA)这一特性来构建自己的研究框架。[①] 这本书提出了合理性与逻辑——尤其是计算机程序和数学公式中能实现的逻辑——之间关系的问题。正如魏岑鲍姆(Weizenbaum, 1976: 11)所论证的:"把计算机引入我们已然是'技术社会'的社会……只是强化和放大了先前的那些压力,这些压力迫使人对自己的社会采取一种日益高度合理性的看法,并为自己塑造一个越来越机械的形象。"魏岑鲍姆认为,应该对给计算机分配的任务设限。这意味着要为计算的部署建立规范限制,因为计算机会影响人在世界上寻找自身定位的欲望。他担心计算机严格遵循工具编程逻辑时形成对同情或人际尊重的伪装,认为这种欺骗将会无意间给双方带来改变的交流性相遇变成一种造成疏离的接触。他(Weizenbaum, 1972: 316)也预见到了对计算机决定提出异议的困难,他写道:

> 人不再为"机器说的话"负责。因此,就不存在对错,不存在正义与否的问题,不存在关于同意或不同意什么的理论。最终,也不存在让人能够质疑"机器说的话"的基础了。

① 在这里,我们可以指出计算系统(尤其是聊天机器人,不管是 ELIZA 还是当代如 Siri 那样的界面)的性别化(更完整的讨论可参看 Marino, 2006)。

重要的是,我们要开始学着从“后端”操作和“前端”界面这两个角度来理解这些新的 AI 系统,因为作为计算系统,它们可能会变得越来越普遍。就像瓦特 1776 年发明的蒸汽机一样,AI 常被视为具有多种潜在应用的通用技术。因此,我们可以预计,AI 的投入使用将对全球经济产生变革性影响。人们已经在日常语境中使用 AI 了,如电子邮件的垃圾邮件过滤功能、媒体的推送系统、导航应用以及支付交易的验证与核实。虽然 AI 显然有对社会和个人产生巨大影响的潜力,但描绘这些影响是一项艰巨的任务,特别是考虑到 AI 领域经常宣扬自己的发现和潜力。① 从监管的角度来看,AI 系统的自主性引发了一系列独特的问题,涉及责任、保险、公平性、风险与安全,甚至创意内容的所有权。这些问题要求我们深入思考透明性与偏见、基础设施决策,以及创造、使用和评判它们所需的各种新数字技能。

我们的社会愈发依赖数字技术,这些技术包含了计算合理性(computational rationalities),进而也包含了算计的合理性(calculative rationalities),并且围绕维持社会学意义上的人类推理能力这一事情提出了诸多挑战(Berry, 2011)。我们对小型软件应用日益依赖,很快就会出现问题,因为它们是自动化、网络化的,并且被接入了更大的软件平台和服务,这使其运作更加复杂。虽然这些系统中有许多最初是为支持或辅助人们进行大量活动、分析和决策的判断而设计的,但它们的运行早已超出用户的理解范围,可同时它们又不可或缺。这样,这些设备通过使人类理性的认知过程卷积短路,通过给予某些逻辑进程工具关系特权,改变了人类理性能力。于是,才有了“我们社会中,个体普遍存在的无力感……人们彼此之间、对自己工

① 在我看来,就技术的变革潜能来说,当前的机器学习和 AI 系统能够把人在当代数字基础设施、数字应用和数字过程中不得不做的“无聊之事”自动化,这一事实比说它们有“创造力”或者能使文化自动化更具说服力。

作的普遍疏离……以及普通人那种生活在庞大系统缝隙中的感觉”(Rosenberg, 1980: 49)。魏岑鲍姆提出的这些说服、欺骗和中介概念,与后来 AI 发展轨迹中普遍存在的问题非常契合。

值得注意的是,发达资本主义社会的经济无政府状态,与理性化和技术相互交织,这种交织倾向于阻碍反思劳动。在此条件下,工具理性的价值被赋予了特权地位,毕竟其体现于合理性概念本身。将算计和理性思考混为一谈,意味着不能被简化为数的事物都是幻想或形而上学的。结果,社会对批判思想的敏感性可能下降,这也体现为个体化潜能的削弱。在当代计算社会,人们有十足的动力采用理性化的行动与思考方式,也就是插入算法的方式。而转向使用 AI,更确切地说是使用机器学习的自动化系统,会进一步推动这个带来疏离感的过程。这表明,能够批判这些系统、避免再次陷入缺乏对系统运行充分依据解释的抽象行为的解释性理解极为重要(Berry, 2021)。正如魏岑鲍姆所说:“即便大型计算机系统,也不一定不可理解。它们可理解性的秘密在于,这些系统是非常稳健的理论模型。”(Rosenberg, 1980: 45)

我们对合理性的构想在多大程度上被重新表述为一种将人的思想理解为算计而非判断的形式,这正是汉娜·阿伦特(Arendt, 1972: 11)联系政治决策和政策所研究的内容。阿伦特的研究显然影响了魏岑鲍姆的思考。[①] 他引用阿伦特的话说:“他们(政策制定者和执行者)不仅聪明,还为(自己)‘合理’……而不下判断感到自豪;他们算计……一种对现实可算计性的从根本上非理性的自信(变成了)决策的主旋律。”(Weizenbaum,

① 魏岑鲍姆提及刘易斯·芒福德、汉娜·阿伦特、雅克·埃吕尔、西奥多·罗斯扎克、肯尼斯·博尔丁,称这些人表达了“对科学技术肆意发展所造就状况的深切关注”(Weizenbaum, 1976: 11)。魏岑鲍姆在《计算机的力量和人的理性》里指出,出版前刘易斯·芒福德“通读了全书”(Weizenbaum, 1976: x)。

1976：14，强调部分为原文所有）。在同一文本的稍后部分，他又阐述道：

> 计算机能够做出司法判决，也能够做出精神疾病诊断。它们抛硬币的方式能比最有耐心的人所采用的方式复杂得多。关键在于，不应该让计算机承担这类任务。在某些情况下，它们甚至能得出“正确的”决定，但这种决定总是且必然建立在没有人愿意接受的基础之上。（Weizenbaum，1976：227）

要联系 ELIZA 来研究这些问题，思考魏岑鲍姆在设计和编写该系统时所关注的语境和一些争论是有帮助的。接下来，我首先要考察魏岑鲍姆开发 ELIZA 的环境，从而理解他是如何构思界面的可供性（affordances）的；其次，思考 ELIZA 与现代用生成式 AI 开发的聊天机器人之间的异同；最后，在结论部分提出关于这种联系的一些反思，并对魏岑鲍姆提出的警告加以引申。

二、ELIZA 和早期的 AI

约瑟夫·魏岑鲍姆于 1923 年 1 月 8 日出生在柏林的一个犹太家庭。他的父亲是皮草技术员，母亲叫亨利埃特。[①] 13 岁时，他随父母逃离纳粹德国，移民到美国并在底特律生活。他后来说自己“是在恰逢合理性的终极流产的成长岁月进入世界的”（Dembart，1977：1）。虽然他最初在韦恩大学学习数学，但 1941 年“他因战争中断了学习，其间在空军气象部门服役”（Sack，2018）。1946 年，他返回大学，并分别于 1948 年和 1950 年获得学士

① 魏岑鲍姆 2008 年 3 月 5 日去世于德国。

和硕士学位。[①] 此后,“他在底特律的韦恩大学参与设计和建造了一台数字计算机”(MIT,2008)。1955 年,他搬到西岸,加入通用电气公司,与 SRI 合作开发了一个名为电子记录机会计(electronic recording machine accounting,ERMA)的银行自动化系统。[②] ERMA 是为帮助美国银行管理此前一直依靠人工处理的海量纸质支票而开发的。在 20 世纪 50 年代初,银行业在本质上处于危机边缘:

> 1943—1952 年,美国的支票使用量翻了一番,从每年签发 40 亿张增长到 80 亿张。银行家估计,到 1955 年,支票数量每年将增加近 10 亿张;到 1960 年,每年签发的支票数量将高达 140 亿张。(Fisher & McKenney, 1993: 44)

ERMA 系统让人们能够利用打印在支票底部的磁编码字体,借助磁墨水字符识别技术实现支票处理自动化。在向媒体做最终演示时,ERMA 还无法完全自动运作,需要一名程序员躲在后面的屋子里帮忙掩盖系统故障。这种用人来替代当前尚未运行或未完成的算法的技巧,被称为“奥兹国的巫师”编程。费舍尔、麦肯尼认为:

> SRI 工程师……以竖大拇指的形式表示系统运行稳定,可继续演示。在设计周期的早期,这种样机形式被视为有益,因其提供了研究和了解用户预期与要求的途径。莫兹利等人认为,这种方式尤其适合探

① 魏岑鲍姆解释说:“促使我投身数学的是……在所有可学的东西里,数学看起来是最容易的。数学就像一场游戏,完全是抽象的。在认为数学最容易这一认识背后,隐藏着与之对应的一种认识,即真实生活是最难的。我从孩童时期起就这么认为了。”(Dembart, 1977)

② SRI 国际是斯坦福大学于 1946 年创立的一家非营利公司,原名为斯坦福研究中心。1970 年,SRI 脱离大学独立,于 1974 年更名为 SRI 国际,且越来越聚焦于国际事务(SRI International, 2016)。

> 索待投入使用系统的设计可能性。ERMA 的展示很顺利,媒体从未质疑(计算机系统)不够理想。(Fisher & McKenney, 1993: 55)

莱茵戈尔德(Rheingold, 1985: 164)认为,“最终将这个过程成功计算机化”是“世界银行系统计算机化的一个里程碑”。不过,在计算过程中,用欺骗代替算法的引入和使用,使其看起来平稳,以及用人力劳动取代计算机这种做法,预见了魏岑鲍姆后来提出的关于自动化的担忧。①

在通用电气任职期间,魏岑鲍姆着手开发一个名为对称列表处理器(symmetric list processor,SLIP)的编程系统,并于 1963 年完成。SLIP 是一组函数,这些函数通过调用例行程序(routines)来执行列表处理,这些例行程序最初由机器代码编写,后来用 Fortan 编程语言编写。魏岑鲍姆本人将其称为“FORTEAN/SLIP 程序”(Weizenbaum, 1963: 524)。在离开通用电气之前,魏岑鲍姆写过一篇名为《怎样使计算机看起来智能》(“How to Make a Computer Appear Intelligent”)的揭秘文章。文章引用马文·明斯基的观点来论证:一种以特定观察者无法理解的方式产生结果的活动,在该观察者看来,在某种程度上是智能的,或者至少有智能的迹象(Weizenbaum, 1962: 24)。在那篇文章中,魏岑鲍姆还描述了一个“五子棋”游戏,该游戏看似智能,但其底层编程实际上并未编入这种智能:

> 根据上述推理,“人工智能”程序的作者能在一段时间内愚弄部分观察者,其成功程度可通过被愚弄的观察者所占百分比乘以他们未察觉的时长来衡量。(Weizenbaum, 1962: 24)

① 考虑到魏岑鲍姆于 1955 年或 1956 年加入这家公司,他是否出席了 ERMA 的发布并不明确。魏岑鲍姆不在费舍尔和麦肯尼所列的主要工程师或贡献者名单上,但他仍有可能听说过那种骗术,即通过在计算机和演示者之间安排一名工程师来假装直接操作计算机。

1993 年,在反思自己工作时,他承认这让他有了“骗子或老千”的名声(Crevier, 1993: 133)。[①] 他觉得:“从某种程度上说,那是我后来开发 ELIZA 的前身,也让我有了骗子或老千的名声。但另一方面,我毫无保留地说了出来。我的想法是创造‘计算机是智能的’这种强烈错觉。在那篇文章里,我花了很大力气解释,幕后没多少东西,机器并没有在思考。我把策略解释得很清楚,任何人都能写出那个程序,我写 ELIZA 时也是如此。”(Creview, 1993: 133)

基于 SLIP 方面的工作,1963 年,他获得了麻省理工学院提供的电子工程助理教授职位。四年间,他就拿到了终身教职,于 1970 年升任计算机科学与工程正教授。[②] 在那里,他用密歇根算法解码器(Michigan algorithm decoder,MAD,1959 年为 IBM7090 和相关计算机开发的一种编程语言与编译器)重写了 SLIP。MAD 基于 ALGOL 语言,但并非 ALGOL 语言。MAD-SLIP 这一术语常被用于指代二者的结合,可实际上,SLIP 和 MAD 始终是分开的,它代表一组可纳入 MAD 程序的列表处理函数。这与长期存在的误解相反,即 ELIZA 最初是用 LISP 而非 MAD-SLIP 编写的。值得注意的是,魏岑鲍姆后来证实,ELIZA 是从他 1962 年编写的那个游戏发展而来

① 我们不禁好奇,魏岑鲍姆是否读过大卫·毛雷尔(D. Maurer)1940 年出版的《大老千》(*The Big Con*)。毛雷尔后来写到,如今,骗子不会再背着一大堆事先写好、涵盖所有可能情况发展的脚本到处跑了,他们什么都不写。不过,他们凭经验就能知道可能出现的情况,也知道他们的台词应对这些情况是完美的。所以,如果内鬼提供了一间办公室,那么就有很多常备的应变方式。骗局的所有参与者都牢记这些应变方式,并且能根据给定的信号采用相应的应变方式(Maurer, 1974: 89; Weil, 2015)。很难想象有比这更好的对 ELIZA 的描述了。有趣的是,魏岑鲍姆也在《计算机的力量和人的理性》(Weizenbaum, 1976: 121)中收录了一小段关于“骗子”的离题讨论。

② 魏岑鲍姆后来先后在哈佛大学教育研究生院、斯坦福大学、柏林工业大学以及德国汉堡大学任职。他还是美国科学促进会会员、纽约科学院和欧洲科学院成员(MIT, 2008)。

的。[①] 克勒维耶认为：

> 20 世纪 60 年代中期，魏岑鲍姆在 MIT 时试图让计算机用英语与人交谈。他指出，像 STUDENT 这样的现有程序应用领域有限，而且描述这些领域的知识与程序结构本身密不可分。魏岑鲍姆还探究了程序通过向用户提问来获取更多信息的有限能力："我很清楚，我们在接下来的几个星期里解决不了这个问题。"魏岑鲍姆告诉我："所以我开始思考替代方案……我把我所有的办法整合起来，做起了 ELIZA 这个项目。"（Crevier，1993：133-134）

ELIZA 试图以一种相对简单的方式模拟人与心理治疗医师可能的对话。它有一个互动界面，用户可通过打字来回答软件所生成的问题。众所周知，很多人对它介入对话的能力印象深刻，觉得这个程序具有比实际更多的理解力与智能。写完 ELIZA 后，魏岑鲍姆承认："我没意识到，极短时间接触一个相对简单的计算机程序，也会让相当正常的人产生强烈的妄想。"（Weizenbaum，1967：7）正如他所解释的："选择这种对话模式，是因为精神科晤谈是为数不多的分类二元自然语言交流（categorized dyadic natural language）的例子之一。在这种交流中，参与双方的其中一方可以自由地摆出对真实世界一无所知的姿态。"（Weizenbaum，1966：42；Boden，1977：108）[②]通过使用简单的模式匹配技巧，魏岑鲍姆让 ELIZA 程序能够把输入

① 魏岑鲍姆（Weizenbaum，1962）所描述的游戏其实是五子棋，而非跳棋。萨缪尔在谈及"机器学习"时暗示了跳棋，但未直接引用（Samuel，1959）。不过，魏岑鲍姆确实参考过香农关于象棋的作品（Shannon，1950；McCarthy，Minsky & Rochester et al.，2006），而且很可能受到萨缪尔 1959 年那篇文章的影响。

② 有趣的是，魏岑鲍姆在一次访谈中称，他"一开始构想 ELIZA 为一名酒保，但后来觉得心理治疗师更有趣"（Crevier，1993：136）。

的英文句子转变为对用户而言看起来有意义的输出。麦考黛克认为：

> 当时，问答机器在 MIT 极为火爆。博布洛（Bobrow）在 MIT 研究 STUDENT；拉斐尔（Raphaelb）在做后来的 SIR；“棒球”程序则是剑桥地区的成果。魏岑鲍姆趁势推波助澜，他连续多日早起，与开发过 COMIT 语言的邻居维克托·英格维（V. Yngve）一起研究模式匹配。（McCorduck，2004：292）

魏岑鲍姆在 ELIZA 程序里探究了许多此类问题。ELIZA 这个名字源于伊莱莎·杜立德，她是萧伯纳 1912 年戏剧《皮格马利翁》（也叫《卖花女》）中的一个工人阶级角色。事实上，魏岑鲍姆后来表示，他是偶然观看了由该戏剧改编、1956 年首演且 1964 年推出电影版的《窈窕淑女》才知晓这个角色的。他为那个后来与 ELIZA 相混淆的人格脚本所起的名字是 DOCTOR。虽然后来这个名字可能一直被视为代表心理治疗师的无性名字，但 ELIZA 延续了这个名字的内涵，将阶级、性别乃至种族与族群的社会结构融入程序中——因为 ELIZA 这个名字包含着萧伯纳戏剧以及 20 世纪 60 年代改编的音乐剧中对女性和社会阶级的再现所蕴含的假设。

ELIZA 会以打字的方式对所有英文输入做出回应，其脚本 DOCTOR 旨在模拟或“戏仿”一位罗杰斯学派心理治疗师。罗杰斯学派心理治疗由心理学家卡尔·罗杰斯于 20 世纪 40 年代创立，采用一种鼓励患者多说话的非指导性语言方法。魏岑鲍姆将这种方法用作构建 ELIZA 回应的模型，ELIZA 的回应大多是重复用户先前对话文本输入中的关键问题。魏岑鲍姆（Weizenbaum，1967：474）将 ELIZA 描述为“后来被称作 DOCTOR 的程序家族中的第一个程序……一个特殊成员。这些程序的家族姓氏也是 ELIZA”。这组程序是在麻省理工学院的一台 IBM 计算机上用 MAD-SLIP 创

建的，发布时只有一个一般性描述（Weizenbaum，1966；Weizenbaum，1967）。在描述程序如何解析句子时，魏岑鲍姆写道：

> 从左至右扫描输入的句子，在关键词词典里查找各个词。若某个词被识别为关键词（除关键词优先级问题外），则只需尝试包含该关键词的分解规则。（Weizenbaum，1966：38）

他接着详细阐述了特定关键词为何需要特殊处理：通常，当一个人使用像“每个人”“总是”或“没有人”这样的全称术语讲话时，实际上指的是某个非常具体的事件或人物。赋予“每个人”比“我”更高的优先级，就能生成“具体说说你想到的是谁”这样的回应（Weizenbaum，1966：39）。

魏岑鲍姆还提供了一个用于说明关键词检测的基本流程图，并描述了一些衡量关键词的具体算法：

> 魏岑鲍姆决定把领域知识放入一个与处理对话的程序模块相分离的模块。他推断，如果用不同的知识模块（他称之为“脚本”）来描述不同种类的知识，程序就能谈论各种各样的话题。今天想谈高级时装而非棒球？只要加载高级时装的软件模块即可！由于程序会像萧伯纳《皮格马利翁》中的伊莱莎·杜立德那样，学着把话说得越来越好，因此魏岑鲍姆将程序命名为《伊莱莎·杜立德》。（魏岑鲍姆后来指出，另一个原因是，“和杜立德小姐一样，它永远不清楚自己是否变得更聪明了”。）（Crevier，1993：134）

三、ELIZA 的社会意义

通过开发 ELIZA 程序和 DOCTOR 脚本,魏岑鲍姆揭示了一点:就计算与人的行为和理性的联系而言,它会带来深刻的影响。魏岑鲍姆试图对这些影响进行探索、提出异议并加以限制:

> 我惊讶地看到,与 DOCTOR 对话的人迅速对计算机产生了深厚的感情,并且毫不犹豫地将其拟人化……这表明,人们以这种方式与计算机对话,就好像它是一个能够与之用亲密的语言恰当、有效地交谈的人。(Weizenbaum, 1976: 6)

事实上,人们倾向于对计算机投入私人情感,这让魏岑鲍姆感到困惑与担忧。他担心机器可能取代人们内心的现实,还担心计算机也许会让"相当正常的人产生强烈的妄想",并且强化那些将合理性与算计联系起来、认为人就是机器的想法。这便是后来所谓的"ELIZA 效应",也就是人认为计算机系统有理解力和智能的倾向。侯世达(Hofstadter, 1995: 167)在 1995 年将其描述为"人的这样一种易感性:人们很容易从计算机串联的字符串(尤其是文字串)中读出远超应有的理解",这个极具说服力的描述也能精确地描述现代的生成式 AI 系统(如 ChatGPT):

> 在这里,我们要记住,魏岑鲍姆版的 DOCTOR 不是在屏幕上一闪而过,而是或多或少被永久地印在纸上。我们可能会把该系统理解为受到文学和心理治疗影响的系统,留下打印记录(如日记)的治疗过程,给人的体验或许和在屏幕上的短暂接触有所不同。(Montfort, 2004)

的确,如今我们更熟悉出现在屏幕上(无论是台式电脑、笔记本电脑、智能手机还是平板电脑)的 ELIZA。但在 20 世纪 60 年代,计算机系统几乎总是使用电传打字机将输出打印在纸上。这种“屏幕的缺失”,是理解当时计算机的物质特性以及使用系统体验的关键。显然,ELIZA 捕捉到了一种思维模式,以及我们与计算机关系中的一种焦虑。魏岑鲍姆后来也通过对自身工作的哲学、社会和政治基础(我们也是基于此理解他的工作的)进行批判性考察,探索了这种思维模式和焦虑。还需指出的是,软件是一种动态且不断变化的媒介,软件程序很少只有一个版本。实际上,在编程过程中常常会出现多个相互矛盾的副本:

> 值得注意的是,ELIZA 存在多个版本。魏岑鲍姆承认自己一直在对 ELIZA 进行改进,因为“最初的 ELIZA 除了生成似是而非的回应外,做其他任何事都很困难”。他在 1967 年写到,后来的版本与旧版本主要有两方面的不同:其一,它包含一个求值器,这个求值器能够接受无限复杂的表达式(程序)并对其求值(执行),当然也能够将求值结果储存起来,以供后续检索和使用;其二,脚本的概念得到了扩展,现在程序可以同时包含三个不同的脚本,能够从磁盘存储单元中无限量供应的脚本里获取新脚本,并且现有的脚本之间还可以相互通信。(Weizenbaum, 1976: 478)

遗憾的是,这些后续的改进缺乏充分的记录,许多源代码仍然处于丢失状态。不过,如今人们已经从 MIT 档案中恢复了 ELIZA 的源代码列表以及

1966 年的一些脚本。[①]

作为计划的一部分,我也参与了重建 ELIZA 代码的工作。我们已经确定,ELIZA 实际上至少有五个主要版本。其中三个版本已被遗忘(并且可能已经丢失),还有一个最终版本只在其他档案的文字资料中被提及。按照年代顺序,我们暂时识别出以下版本:

1. ELIZA 1965a:仅以“.”和“,”分隔句子(证据源于 MIT 档案流程图)。

2. ELIZA 1965b:用“.”“,”和“但是”来分隔句子,缺乏 NEWKEY 函数,包含未被记录过的 CHANGE 函数以及硬编码信息(这是从 MIT 档案中恢复的版本)。

3. ELIZA 1966 CACM:包括 NEWKEY 函数、关键词堆栈和“but”分隔符。

4. ELIZA 1967:添加了复杂的脚本处理,证据不仅来自魏岑鲍姆和泰勒的描述,也见于档案中现存的 ARITHM、F29、FIGURE 脚本。

5. ELIZA 1968+:对这个版本的计划描述见于泰勒。不过,脚本 SPACKS、INTRVW 和 FVP1 也证明,在更为复杂的编程中,通过瓦尔特·丹尼尔(W. Daniel)开发的脚本文本,使用过这个版本。

---

① 除 DOCTOR 外,MIT 档案中的脚本如下:ARITHM(可进行对话中的数学计算,求值并返回结果)、F29(似乎是 FIGURE 数学定义脚本的早期版本)、FIGURE(像是 F29 加上更多定义和回应后的拓展版本)、GIRL(简单演示脚本)、NEWENG(用于讨论新英格兰各州)。有脚本输出但无脚本本身的有:ELEVTR(讨论电梯物理学)、POLETA(讨论“竿与谷仓悖论”)、SPACKS(讨论 MIT 人文系引用的贝利·斯帕克斯的一句诗)、SYNCTA(讨论时间同步脚本的名称)。档案中还提及一些既无脚本也无脚本输出的脚本:ANTIPR、INTRVW(据描述是为四向量研究做准备的初步采访)、FVP1、FVQUIZ、CANVEC、FRANCE、FORVEC、MIT、ORTH1、PHOTON、RLPOLN、STATES、QMPROB、WATSNU、XYPOLN。相关重建工作可参看 https://wg.criticalcodestudies.com/index.php?p=/discus-sion/108/the-original-eliza-in-mad-slip-2022-code-critique。

魏岑鲍姆认为,ELIZA 开启了人类依赖和信任计算机的先河,而未来这样的例子还会越来越多。然而,这只是更大问题的一个元素而已,这个问题就是:

> 科学向人类许诺了权力。但就像在人类被权力的许诺诱惑时经常发生的那样,他事先就开始并且在整个过程中都在付出代价,而实际付出的代价,是奴役和无能。权力如果不是选择的权力,那它就什么也不是。工具理性可以做决定,但决定和选择完全不一样。(Weizenbaum, 1976: 259)

当时,人们围绕 ELIZA 和 PARRY 展开了激烈的争论——PARRY 是由肯尼斯·科尔比(K. Colby)于 1972 年设计的一个竞品系统(关于二者的结合,参看 Cerf, 1973)。科尔比试图创造一个具有偏执型精神分裂症特征的聊天机器人,这在当时引发了巨大的争议,并引起了人们对这类软件意义的广泛关注,这些争论也继续影响着当代围绕聊天机器人和 AI 的讨论。

ELIZA 提出的问题在大众文化中得到了更广泛的响应。例如,“火象剧场”(The Firesign Theater)在 1971 年的专辑《我认为我们都是这辆巴士上的笨蛋》(*I Think We're All Bozos on This Bus*)中虚构了它的各种可能性。在这张专辑中,主人公克莱姆去“未来博览会”参观,在那里,他与一台用模糊、听起来积极且与提出的问题几乎无关的回复来回答的计算机交谈。当克莱姆(他的名字被博览会的人格化系统错听成“阿克莱姆”)有机会提问时,他决定黑掉它,通过说“这是工作人员在说话。你好……”进入系统的维护模式;那台名叫 DOCTOR MEMORY 的计算机用“系统状态:运行时间”和运行时长来响应他。接着,克莱姆试图通过提出无法理解,有时甚至

无法解析的问题来使系统混乱,从而使系统崩溃。最后,克莱姆问:“为什么粥鸟要在空中下蛋?”这使得系统停止服务并关闭。这张专辑中出现了一个类似于 ELIZA 的系统,这让许多懂行的人好奇,这样内行的技术行话怎么会出现在一张喜剧专辑上。这一直是一个谜,直到 2015 年,“火象剧场”最初的成员之一菲利普·普洛克特在 Quora 上用一条评论解释说:“我(那张专辑里)所有的计算机语言,都是从我在洛杉矶工作博览会上找到的一张人们与 ELIZA 心理治疗师程序互动的打印资料上学来的。”(Proctor, 2015)

苹果的 iPhone 的 AI 聊天机器人 Siri 后来也提到了这个“大众文化时刻”,它会用“你好阿克莱姆,我能为你执行什么功能?”来回答“这是工作人员在说话。你好……”①至于这个段子是怎么被放进 Siri 的,根据普洛克特的说法,是“史蒂夫·乔布斯在皮克斯一部普洛克特担任配音演员的电影放映会上遇到他,告诉他自己是‘火象的忠实粉丝’,并要求把这个桥段加入 Siri”(Tannenbaum, 2015)。② 如这里所展示的,在这些例子中,ELIZA 的影响和它提出的问题,从技术领域转移到了文化领域,并加剧了智能机器出现给公众带来的不安。③

在魏岑鲍姆 1966 年的文章发表后不久,伯尼·科赛尔(B. Cosell)就用 LISP 编写了一个 ELIZA 程序。这个程序以魏岑鲍姆公布的算法为基础,不过重要的是,科赛尔并未见过实际的 MAD-SLIP 源代码。LISP 很快成为 AI 的主要语言,而魏岑鲍姆的 DOCTOR 脚本格式正是 LISP 的符号表达

① 可惜的是,这个彩蛋在过去几年间的某个时候被移除了。

② 最近的对话界面(也被称为可听界面)复制了 ELIZA 的诸多特征。这一情况在约翰·凯莱(J. Caylay)的《聆听者》(*The Listeners*)中也有体现。它是在扬声器或通过扬声器访问的用户与 Amazon Echo 之间进行的语言表演、装置,是由亚马逊分发的第三方应用。

③ 最近,ELIZA 又出现在阿德里安·柴可夫斯基(A. Tchaikowsky)2015 年的科幻小说《时间之子》(*Children of Time*)里了。

方式。此外,ELIZA 的原始版本仅在学术圈小范围流传,这些因素很可能加深了那个长期存在的误解,即 ELIZA 一开始就是用 LISP 编写的。需要注意的是,科赛尔的代码也未曾广泛公布,而且该代码直到 2013 年才被恢复。科赛尔最著名的版本是 1966 年的那一版,GNU 的 EMACS 版和杰夫·施拉格(J. Shrager)的 BASIC 版都是基于此版本编写的。后者最初发布于 1973 年,1977 年由《创意计算杂志》(*Creative Computing magazine*)再版。① 施拉格的 ELIZA 很短,能够快速手动重新键入,并且可以在任何支持 BASIC 语言(当时这种情况很常见)的个人计算机上运行(North, 1977)。② 如今,多个系统上仍然存在类似的程序,包括手机版和能在网页浏览器上运行的 Java 版。这些版本就运行的程序语言而言稍有不同,并且仅在长度上存在差异——科赛尔的 LISP 版包含 2500 行代码,施拉格的 BASIC 程序仅有 250 行,其他版本则介于两者之间。③

魏岑鲍姆后来在《计算机的力量和人的理性》中继续阐述操纵和误认的危险。在该书中,他认为计算机会给人类用户带来无法预见的负面影响。魏岑鲍姆早期有将人的工作自动化(如用计算机器取代银行职员)的经验,这影响了他日后对计算机给社会带来挑战的理解。魏岑鲍姆尤其担心人的理性会被有效的程序所取代。对他而言,有效的程序是一种操作,这种操作能让我们在形式数学里定义一个问题域,使该问题域能被机器进行后续计算。而这种思想的形式化又为将人类生活的各方面贬低为机器程序(如把

① 理查德·斯托曼(R. Stallman)一直被视作 GNU 的 EMACS 版 ELIZA 的作者,但斯托曼告诉我,他并未编写过这个程序(个人通信,2022 年 11 月 13 日)。

② DOCTOR 是伯尼·科赛尔依照魏岑鲍姆在期刊文章中的描述编写的 LISP 版 ELIZA。伯尼·科赛尔是 BBN(Bolt Beranek and Newman)科技公司程序员团队成员之一。1972 年,该团队制造出两台用作 ARPANET 通信路由的专用计算机,名为接口信息处理机(Interface Message Processor,IMP)。1972 年 9 月,文顿·瑟夫(V. Cerf)录制了 DOCTOR 和 PARRY 的一次对话,并于 1973 年将结果记录为征求评论稿(Cerf, 1973)。

③ 我要感谢 2016 年与 2022 年的批判代码研究工作团队(Critical Code Studies Working Group),他们对 ELIZA 展开的丰富且详尽的讨论,极大地启发了我对 ELIZA 和 DOCTOR 的思考。

计算当作思考)创造了条件,引出大多数人的这一问题:

> 根本不理解计算机。所以,除非他们有很强的怀疑能力……否则就只能运用自己能想到的唯一一个类比,即按照自己的思考能力为模型,来解释计算机的智力壮举。(Weizenbaum, 1976: 10)

也就是说,存在一个共同假设:计算机对行为的抽象有助于解释机器的运作。在当代计算社会中,人们有充分的动力采用理性化、插入算法的行动和思维方式,自魏岑鲍姆写下相关文字以来,这种动力有增无减。例如,引入可测量的性能指标和输出标准,以及以计算为前提的监控和监视。我们还发现,在资本主义社会中,相较于互动(对话),常规工作反而增多了。同样严重的是算法系统缺乏合法性,部分原因在于,这些系统在加剧与民主权威相关的结构问题(包括它们在通信和媒体系统中日益广泛的部署)的同时,自身仍然不透明,它们不断增强的影响力表明,它们有可能产生系统性危害,造成危险的系统故障(Berry, 2014)。也许,只要稍微回顾一下2008—2011 年的美国次贷危机,就能发现与理性化和计算相结合的"算计式理性"(calculative reason)与以利润为导向的公司和个人会形成多么令人头疼的组合。魏岑鲍姆敏锐地察觉到了这些问题,并试图阐明其中的重要议题,他评论道:"我们的社会对计算机系统(最初,我们只想让它们'帮助'人们进行分析和决策)的依赖日益加深,早已超出用户的理解范围,并且对用户来说变得不可或缺(……这)是一个非常严重的发展。"(Weizenbaum, 1976: 236)

在这里,魏岑鲍姆所指的是算法化带来的物化与形式化,以及我们构建系统的方式。在创建系统时,人们不再确切知晓或理解系统的运作及决策方式。这就产生了一种危险:因害怕未知的后果,我们不愿修正系统。这进

而引出了增强技术所体现和支持的人的主体间历史概念的创造与维持问题。魏岑鲍姆 1976 年的评论就触及了这一点,他说:“《纽约时报》已经开始建立一个实时数据库……这离让系统来判定什么算事实,进而直接宣布其他所有知识、所有记忆不合法还有多远呢?”(Weizenbaum, 1976: 238)在复杂系统、庞大的数据集合与快速流动的数据流之间进行协调的能力,为社会将“规模”作为一个新的智识视域来运用提供了条件。在业界和知识界的研究中可以看到,人们越来越多地运用计算系统对复杂的“大数据”现象进行抽象、简化和可视化处理,并且倾向于使用过于简单化的因果和统计模型来理解复杂的社会现象,“社会物理学”概念就是一个例证(Pentland, 2015)。魏岑鲍姆可能会评论说,这种“技术、政治和社会的不可避免性”的混合,“对良知来说是一种强大的镇静剂”,它不但赋予计算决定的权力,还赋予其行动的权力(Weizenbaum, 1976: 241)。

魏岑鲍姆依据他和 ELIZA 及其他系统打交道的经验论证称,我们必须对人工智能及其技术发展持批判态度,如此才能理解当下的处境,并且防止其失控:

> 我不太清楚计算机科学及其分支学科(AI)是否尤其偏爱委婉。我们如此夸张又欣然地谈论能理解、能观看、能决定、能下判断等的计算机系统,却不承认自己在这些概念上的浅薄与极度的天真。而且,在这样的谈论过程中,我们麻痹了评估自身工作质量的能力,更重要的是,麻痹了识别和认知其最终用途的能力。(Weizenbaum, 1987: 44)

四、现代的迭代

魏岑鲍姆指出,人的思想和劳动自动化存在危险。这要求我们以一种拒绝忽视和平息现象层面的矛盾及矛盾主张、努力把握主体动态时刻(dynamic moment)的方式来解释。这种解释方式要始终为批判反思理解算法的历史和传统留出空间,接受传统意义结构的重要性,同时力求避免将其理想化。也就是说,在这些计算社会中,传统可能继续表现为基于欺骗和扭曲的互动,而这种互动往往会被不加反思地转换为算法形式,用魏岑鲍姆的话来说:

> 我们所讨论的各种系统的程序都具备一些非常重要的特征:从某种意义上说,它们都很简单;它们都会扭曲和滥用语言;在否认规范内容的同时,它们也都支持一种基于专业知识的威权主义。(Weizenbaum, 1976: 248)

现在,为进一步强调这点,我想简要考察一个例子,它将说明 ELIZA 的界面可能性与我们对计算机功能的预期为何可能以极其反人类的方式互动。在这个例子里,社会冲突内嵌于算法机器,劳动通过界面被转化为一种商品。这个例子就是用于隐藏社会劳动,使工人隐藏在网页表单、聊天机器人和应用程序接口(这些能让雇主将自己视为创新技术的建造者,而非不关心工作条件的老板)背后的计算(Irani & Silberman, 2013: 613)。在此语境下,算法造就的"自由"和"创造"可能实际上只是掩盖了控制和管理系统而已。

为理解社会冲突如何隐藏于算法形式之中,研究在新计算技术条件下

才成为可能的现代生成式 AI 系统(或聊天机器人)是有益的。从某种意义上讲,这是“大脑之间的合作”。斯蒂格勒(Stiegler, 2010: 47)指出,这种合作“通过语法化系统产生,该系统使神经系统活动最高水平上的所有质性任务的无产阶级化成为可能”。关键是,这是对劳动力的实时抽象,将劳动力抽象为一种潜能,一种像电力或水一样按需持续供应的“劳动力流”。经由聊天机器人和应用的调节,这种抽象造就了一种高度异化的劳动力。从某种意义上说,这种劳动力实现了控制论系统,该系统试图在紧密联系的反馈回路中让机器与人紧密耦合。这也描绘了魏岑鲍姆在《计算机的力量和人的理性》中所警告的威胁。

这方面的范例当属 OpenAI 的 ChatGPT。该公司首席执行官称 ChatGPT 能改变资本主义:“技术的改进及其对社会的积极影响将持续指数级增长。”(Konrad & Cai, 2023)ChatGPT 是基于生成式预训练 Transformer 模型[大语言模型(LLM)的一种]的聊天机器人,LMM 是基于真实对话、文本以及像 Reddit 那样的平台上的文章训练而成的。它让当代观察者感到惊奇,因为通过输入和问题来提示系统,它似乎将书面文本的生产机械化了。同时,系统的力量实际上隐藏在聊天机器人的“界面”之后。这种既通过软件汇集人类文化生产,又将其作为计算系统储备资源的想法,体现了魏岑鲍姆有先见之明地警示过的计算社会中流行的控制论思想。从很多方面来看,这使人类活动变得离散。不过,这也是人通过用于协调更普遍文化生产的计算层的非人化。它还表明界面如何将表面之下进行的社会劳动物化,从而使机器能够真正将数百万人类写作和对话的痕迹据为己有,而这一切都是人们没有意识到且未曾同意的。

这也开启了按需求进行管理、控制、监控、分解与再集合的自动化文化生产的可能性。资本主义公司将其操作化后,就有造成社会休克的危险,因为新的技术实践往往能避开劳动法与劳动保护的监管,还能带来巨大利润。

这可能扭曲更广泛的经济,这是缺乏政府监管和工会等劳动组织监督的结果。或许我们还会发现,未来的社会冲突与这种自动化系统的扩大使用有关。实际上,它们营造出一种算法拜物教的感觉,劳动被算法掩盖,这意味着通过面板界面和编程代码互动的程序员或用户无须承认劳动的社会性。对界面另一边的工人来说,对能动性、体力劳动、情感控制和自我规训的要求可能造成严重的心理紧张(在失范与宿命论之间切换)。消费者或客户的一条差评就可能立即导致失业,而且资方几乎不会给出解雇原因。这是否会为必要性政治和社会学反身性创造条件还有待观察,但它肯定是传统工人阶级之外未来劳动祛魅的一个可能来源。我们在当前流行的基于文本界面的聊天机器人中也能看到类似的例子。和 ELIZA 一样,这些聊天机器人也是部分自动化、部分由人来管理客户的询问和问题(Marino, 2006)。不用说,ELIZA 启发创造认知工厂的计算系统这一想法会让魏岑鲍姆感到恐惧。

通过使用这种新范式,运用生成式 AI 的系统指出了新一类知识—权力化界面的各种可能性。在此,机器被界面模糊,它可能是人类劳动、聊天机器人或者其他形式的 AI。从界面用户的视角看,没有任何差别。这个程序将关系物化,在用户与底层过程之间创造出一种命令—执行关系。对手段的模糊(不论它可能编码何种剥削关系)以及在计算系统中对真实人类进行理性化和抽象所带来的后果,也让人联想到魏岑鲍姆 1976 年的评论:

> 当个体不被当作完整的人来对待时,就被非人化了。各种形式的人类和社会工程……在绕过人的语境,尤其是绕开那些赋予人的语言以真实意义的语境时,就在做这样的事。(Weizenbaum, 1976: 266)

五、关于生成式 AI 意义的反思

通过考察先前有关早期计算的论证和讨论,我们找到了一种反思当下困境与问题的有用方式,这些困境与问题出现在我们急于将社会合理化和计算机化之时。身为计算科学家,魏岑鲍姆乐于反思自己的实践及其工作的意义:

> 这应该教会我们什么呢?特别是就至少保住甚至增加人在人类事务中的选择这一问题而言。当然是,与其说构建可靠的计算机软件依靠计算机科学的研究成果,不如说是对人类状况更深入的理论理解。(Rosenberg, 1980:103)

虽然 ELIZA 放在今天或许显得原始,但它仍反映出诸多设计决策,这些决策在我们如今使用的系统中依旧有所体现,比如人机如何互动,以及应在多大程度上允许计算管理我们的生活与心智等。它还表明,早期软件的发展脉络对于理解我们在何种程度上必须始终对过去的假设以及技术可能且应该如何发展的想法保持质疑非常关键。虽然如今我们不会把 ELIZA 视为 AI(即便过去我们曾有过这种错误认知),但我们也应认识到,类似的工具合理性仍然内置于机器学习和当今的 AI 系统之中。

今天的计算机科学环境与魏岑鲍姆当时的环境有实质性不同。他在学院环境下工作,虽得到美国国防部 MAC 计划的项目资助,却从实验研究视角探索技术(如 ELIZA)。如今,AI 日益被资本主义需求所涵盖(Berry, 2021; McQuillan, 2023)。多数 AI 的发展趋向计算的历史规范,包括强调技术层面将控制与执行分开的能力。计算的批判史显示,这种方法常被用

于与资本主义相符的进程,其中隐含一套假设,如在构建社会关系时对市场优先性的先验假设。其次——在我看来与之不无关联——当前理解 AI 的方式倾向于促使人们对其运作进行形而上学或形式主义的辩护与解释。这可能使思想的数学化成为一种价值,在这种价值观念下,思想的形式化不仅是思考 AI 的一种方式,而且是唯一可效仿的方式。这可能催生一种倾向于理念论,却不关注谁拥有和控制新“认知手段”的计算理论,就如布莱兹·帕斯卡尔在 17 世纪评论的那样:

> 数学的心智(esprit de geométrie,几何学精神)和直觉的心智(esprit de finesse,敏感性精神)是存在差别的。数学家不讲直觉,原因在于他们看不到眼前的事物。他们习惯了明确、显而易见的数学原理,习惯在正确理解和处理原理之前就得出结论,所以在需要直觉的问题中就会迷失。在这类问题里,原理不能像之前那样被处理……这些事物极为微妙且众多,需要一种非常微妙和精确的性情去感受它们,然后依据这种感知做出准确、正确的判断。最常见的情况是,我们无法像在数学中那样从逻辑上对其进行证明,因为我们认识原理的方式不同,那样做会是一项无尽的任务。“真”必须直接去看,一眼便知,至少在一定程度上不需要通过推理过程。所以,数学家很少按直觉行事,凭直觉行事的人也很少是数学家。因为数学家想用数学的方式处理直觉方面的事物,他们荒谬地想从定义开始,再按照原理……倒不是说心智不会这么做,而是说,心智会沉默、自然而简单地这样做。(Pascal, 1995: 151,译文有改动)

在不否认人文学科或社会科学对 AI 或机器学习的研究方法尚不成熟的情况下,我想魏岑鲍姆会认同,AI 思考中的严重问题确切地说正是这样

一个假设:数学化或形式化是理解的前提。毕竟,魏岑鲍姆本人解释过:

> 他们一开始可能只是相信,自己采用的运算不过是在描述所处理现象时一种便捷的速记方式。但随着他们用从外部语境借来的基本要素构建出越来越庞大的概念框架,为这些框架命名,并且像操控更复杂思想体系的元素一样操控它们,这些框架就不再仅仅是描述方式了,而是像马斯洛的锤子一样,成为他们世界观的规定。(Weizenbaum,1976:102)

随着以 ChatGPT 这类系统为代表的机器学习不断进步,人类思维能力被毁灭这一危险(可能会被误解为只是 AI 偶然的副作用)已随之而来了。不过,对于任何技术而言,在创造技术、应用技术以及围绕技术在社会中的使用进行限制和管制方面,人类总有选择。魏岑鲍姆富有先见之明地认识到,无知、无意义的机器可能会对我们的理解能力、主体间交流能力以及基本人性构成挑战。他发现人类对彼此存在自然的同理心,也明白这种同理心可能会被错误地导向计算机,还认识到对于技术主义者来说,利用这种同理心误导或欺骗人们以谋取利益、煽动或说服他们,乃至破坏民主本身具有巨大的诱惑。我们正生活在 ELIZA 所指向,而魏岑鲍姆试图警示的未来。现在,听从他的警告,明确可计算事物的限度是我们的责任。①

首先,魏岑鲍姆开发 ELIZA 的初衷是探索计算机与人类交互的可能性。在当时的技术环境下,这是一项极具开创性的工作。他的贡献不仅在于创造了一个可模拟对话的程序,更在于引发了人们对人工智能本质的深

① 在这里,我对“不可计算的东西”的理解,是从哪类问题适合计算的批判意义出发的,而非依据关于什么能或不能被算法再现的更具技术性的定义。在此语境下,不可计算的东西就是被社会定义为不适合计算的那些问题。

入思考。

在生成式 AI(如 ChatGPT)蓬勃发展的当下,魏岑鲍姆批判性更强的著作显得格外重要。他曾警示过过度依赖人工智能可能产生的风险,像人机关系失衡以及对人类思维方式的潜在影响等。ELIZA 虽是较为初级的程序,却已展现出计算机模拟人类对话的潜力,当代的机器学习和 AI 技术更是将这种潜力发挥到了新高度。

ELIZA 相关讨论为当下机器学习与 AI 的争论提供了独特视角。一方面,ELIZA 使人们认识到计算机在一定程度上能模仿人类语言行为,这为机器学习奠定了早期理论与实践基础;另一方面,人们也意识到仅追求语言相似性并不代表计算机具备人类智能。比如,ELIZA 的回应常基于预设模式和算法,缺乏真正的理解与情感。

当代,生成式 AI 面临诸多挑战与争议,如数据隐私、算法偏见、对就业市场的冲击等。魏岑鲍姆的思想能引导我们更谨慎对待这些发展。我们不能只陶醉于 AI 带来的便利和高效中,还得关注其背后可能潜藏的风险。同时,对于 AI 的发展方向,我们应更注重让其与人类社会更好地融合,而非单纯追求技术超越。

总之,考察魏岑鲍姆对 AI 的贡献并讨论 ELIZA,有助于我们更好地理解当前机器学习和 AI 发展中的关键问题,从而在这个快速发展的领域保持清醒头脑,引导 AI 朝着对人类有益的方向发展。

参考文献

Arendt, H. 1972, *Crises of the Republic*, [S.l.]: Harcourt Brace Jovanovich Publishing.

Berry, D. 2011, *The Philosophy of Software: Code and Mediation in the Digital Age*, [S.l.]: Palgrave.

Berry, D. 2014, *Critical Theory and the Digital*, [S.l.]: Bloomsbury.

Berry, D. 2021, "Explanatory Publics: Explainability and Democratic Thought." in B. Balaskas & C. Rito eds., *Fabricating Publics: The Dissemination of Culture in the Post-Truth Era*, [S. l.]: Open Humanities Press.

Boden, M. 1977, *Artificial Intelligence and Natural Man*, Cambridge, MA.: MIT Press.

Cerf, V. 1973, "RFC 439-PARRY Encounters the DOCTOR." Faqs. org. http://www. faqs. org/rfcs/rfc439. html.

Crevier, D. 1993, *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York: Basic Books.

Dembart, L. 1977, "Experts Argue Whether Computers Could Reason, and If They Should." https://www. nytimes. com/1977/05/08/archives/experts-argue-whether-computers-could-reason-and-if-they-should. html.

Fisher, A. & J. McKenney 1993, "The Development of the ERMA Banking System: Lessons from History." *IEEE Annals of the History of Computing* 15(1).

Hofstadter, D. 1995, *Fluid Concepts and Creative Analogies*, [S. l.]: Perseus Books.

Irani, L. & M. Silberman 2013, "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." https://doi. org/10. 1145/2470654. 2470742.

Konrad, A. & K. Cai 2023, "Exclusive Interview: OpenAI's Sam Altman Talks ChatGPT and How Artificial General Intelligence Can 'Break Capitalism'." https://www. forbes. com/sites/alexkonrad/2023/02/03/exclusive-openai-sam-altman-chatgpt-agi-google-search.

Marino, M. 2006, *I, Chatbot: The Gender and Race Performativity of Conversational Agents*, Ph. D thesis of University of California.

Maurer, D. 1974, *The American Confidence Man*, [S. l.]: Charles C. Thomas Publisher.

McCarthy, J., M. Minsky & N. Rochester et al. 2006, "A Proposal for the Dartmouth Summer Rssearch Project on Artificial Intelligence." https://doi. org/10. 1609/aimag. v27i4. 1904.

McCorduck, P. 2004, *Machines Who Think*, [S. l.]: A. K. Peters.

McQuillan, D. 2023, "Predicted Benefits, Proven Harms: How AI's Algorithmic Violence

Emerged from Our Own Social Matrix." https://doi.org/10.51428/tsr.ekpj9730.

MIT 2008, "Joseph Weizenbaum, Professor Emeritus of Computer Science, 85." https://news.mit.edu/2008/obit-weizenbaum-0310.

Montfort, N. 2004, "Continuous Paper: The Early Materiality and Workings of Electronic Literature." http://nickm.com/writing/essays/continuous_paper_mla.html.

North, S. 1977, "Eliza." *Creative Computing* [s.l.].

Pascal, B. 1995, *Pensées and Other Writings*, Oxford: Oxford University Press.

Pentland, A. 2015, *Social Physics: How Social Networks Can Make Us Smarter*, London: Penguin.

Proctor, P. 2015, "What Are the Best Siri Replies?" https://www.quora.com/log/revision/82373474.

Rheingold, H. 1985, *Tools for Thought: The History and Future of Mind-Expanding Technology*, Cambridge, MA.: MIT Press.

Rosenberg, R. 1980, *Incomprehensible Computer Systems: Knowledge Without Wisdom*, Ph. D thesis of MIT.

Sack, H. 2018, "Joseph Weizenbaum and His Famous Eliza." http://scihi.org/joseph-weizenbaum-eliza.

Samuel, A. 1959, "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development* 3(3).

Shannon, C. 1950, "XXll. Programming a Computer for Playing Chess." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41(314).

SRI International 2016, "About Us." https://www.sri.com/about.

Stiegler, B. 2010, *For a New Critique of Political Economy*, London: Polity.

Tannenbaum, S. 2015, "With One of Its Easter Eggs, SIRI Evokes the Firesign Theatre, Hacker Culture, a 1960s Chatbot, and Steve Jobs." https://medium.com/@stannenb/with-one-of-its-easter-eggs-siri-evokes-the-firesign-theater-and-steve-jobs-86ea5b4874d3#.rn9smlfgm.

Weil, P. 2015, "Seriously Writing SIRI." doi:10.20415/hyp/011.e05.

Weizenbaum, J. 1962, "How to Make a Computer Appear Intelligent." *Datamation* 8(2).

Weizenbaum, J. 1963, "Symmetric List Processor." *Communications of the ACM* 6(9).

Weizenbaum, J. 1966, "ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine." *Communications of the ACM* 9.

Weizenbaum, J. 1967, "Contextual Understanding by Computers." *Communications of the ACM* 10(8).

Weizenbaum, J. 1972, "On the Impact of the Computer on Society." *Science* 176.

Weizenbaum, J. 1976, *Computer Power and Human Reason: From Judgement to Calculation*, [S. l.]: W. H. Freeman and Company.

Weizenbaum, J. 1987, "Not Without Us." *ETC: A Review of General Semantics* 44(1).

从约瑟夫·魏岑鲍姆到 ChatGPT*

——与令人倾倒的 AI 技术的批判交锋

克里斯蒂安娜·弗洛伊德　著**

王立秋　译***

摘要:本文从使用的角度对 AI 系统进行思考。聚焦于对话聊天机器人,从魏岑鲍姆的 ELIZA 谈起,概述到 ChatGPT 的重大科学进步。文中的主要讨论基于作者为探索 ChatGPT 这一知识资源而进行的数次实验—反思循环。分析部分从准确性、结构、语境、视角和偏见等角度思考 ChatGPT 的回应。批判性评估始于这样一个观察:ChatGPT 会在未明确自身能力范围的情况下,给出清晰准确结果与任意错误信息的混合。这引出了系统的关键问题,即它如何与"真"相抗衡,这又涉及用基于文本相关性的概率替代品取代"真"的理念。在负责任的使用中,ChatGPT 之类的系统必须融入人类学习文化。这一过程的框架应包括坚持真实性、提升人类能力的动力以及强化共同体内的责任结构。

关键词:聊天机器人　人文主义　以用户为中心的设计　人工智能批判

[在]人工智能[领域……],人们制造机器,是想让它们做出惊人之举,这些举动往往会令最有经验的观察者为之倾倒。然而,一旦特定程序被揭露,一旦其内部运作得到解释[……]它的魔力就会瞬间消失;它不过是一堆程序代码[……]观察者会心想"这我也能写"。抱着

---

* 本文原题为"From Joseph Weizenbaum to ChatGPT",原刊于 *Weizenbaum Journal of the Digital Society* 3(3),由作者以 CC-BY 协议许可翻译发表。

** 克里斯蒂安娜·弗洛伊德(C. Floyd),德国汉堡大学荣休教授。

*** 王立秋,哈尔滨工程大学人文社会科学学院。

这种想法,他就会把这个程序从“智能”的架子挪到为好奇者保留的架子上。(Weizenbaum, 1966)

作为 AI 语言模型,我不像人类那样拥有知识,也不具备自我反思或内省的能力。(ChatGPT, April 3, 2023)

一、引言

2023 年 1 月 11 日,我有幸在魏岑鲍姆研究所为纪念魏岑鲍姆诞生一百周年的活动做开幕演说。当时,在场之人无一人预见,短短几周后,随着 ChatGPT 公开发布,人工智能(AI)会热度惊人。

作为聊天机器人,ChatGPT 是约瑟夫·魏岑鲍姆所创造的、影响深远的程序 ELIZA 的直系后代。ELIZA 首次让人们能够通过电传打字机进行所谓的人机“对话”,而且(按照不同脚本)可定制,能扮演多种对话角色(Weizenbaum, 1966)。其 DOCTOR 版本加载的脚本让它能像罗杰斯学派心理治疗师那样行动,该版本被广泛使用且引发广泛讨论。开发 DOCTOR 的经历,也让魏岑鲍姆从一名充满热情的 AI 研究者,转变为最著名的 AI 批评者之一。

让魏岑鲍姆感到震惊的是:(1)执业心理治疗师认为,可以加强 DOCTOR 以实现心理治疗自动化;(2)用户与 DOCTOR 互动时仿佛它是一个人,会向它透露自己的秘密,还将人的品质赋予技术系统;(3)很多人认为可将该程序普遍化,为理解整体自然语言提供基础。

在本文开头所引用的魏岑鲍姆于 1966 年发表的关于 ELIZA 的文章里,他选用“令人倾倒”一词描述 DOCTOR 对许多人的影响,还鼓励人们去仔细探究其“内部运作”。如此一来,人们就会发现“魔力瞬间消失”,有经验的观察者则会意识到自己也能编写这样的程序。

不过,正如 ChatGPT 所展示的那样,所有聊天机器人的前身与现代 AI 语言模型之间存在巨大的技术差距。我们有足够多的理由对其“倾心”,以至于许多人觉得自己作为人的身份被动摇,并且对新技术的影响感到恐惧。我见过不少这样的人,他们记住“人工智能”这个名字只是因为它突然变得无处不在。与 ELIZA 相比,我们对于如何充分理解这类系统的“内部运作”从而让“魔力瞬间消失”还毫无头绪。

甚至“最有经验的观察者”也不敢声称仅凭自己就能写出这样的程序。这个系统极为复杂,它有着复杂精细的内部架构,与其他技术有着错综复杂的联系,而且依赖于环境,尤其是用于训练的文本语料库。并且,基于接触过的文本数据,它的行为会随时间发生变化。ChatGPT 所包含的深度学习装置,让人们难以理解其机制,难以对其进行控制,也难以确保使用者不会受到伤害。实际上,在向公众开放该系统几个月后,其开发者就迫切呼吁对它进行监管。

我不是 AI 专家,但我有幸在 1968 年就见识过 AI。当时我是斯坦福人工智能实验室的研究助理,在那里,我目睹了一批令人惊奇的机器,它们被制造出来就是为了做出惊人之举,其中许多是后来发展成熟、在日常生活各种语境下投入使用的 AI 技术原型。其中之一就是 DOCTOR,我见证了它对信任它的用户的影响。

我很早就听闻一些顶尖 AI 专家的惊人言论,他们将人等同于机器。我还记得有一位专家在一次演讲中向听众挑衅:“当然,人像机器,在场有谁觉得受到威胁了吗?”正是“威胁”这一元素,让他的话令人难忘。我的 AI 界朋友大多与这类主张保持距离,也有人站出来反对,其中最著名的当属约瑟夫·魏岑鲍姆。自 20 世纪 70 年代早期在斯坦福与他结识后,我就与他保持着长久的友谊。他的开创性著作《计算机的力量和人的理性》(1976)为该话题的讨论设定了标准。反对将人等同于机器,也成了我自身专业工

作的一个支柱(Floyd, 1986),但我仍对 AI 取得的巨大技术成就满怀钦佩。

本文将从使用的角度对 AI 系统进行思考。AI 界不喜欢谈论 AI 系统的使用,而倾向于从以技术为中心的视角来讨论影响,仿佛技术可以自行行动。令我失望的是,我发现,就连斯坦福大学的以人为中心的人工智能中心(Center for Human-Centered Artificial Intelligence at Stanford University)也遵循这种思维方式。对他们而言,“以人为中心”意味着在探索人与 AI 的相似与不同的理论基础上,推动改进人与 AI 共生关系的研究。

我从另一种意义上的“以人为中心”这一角度切入。我想要思考的是:作为工具或者媒介,AI 系统如何才能适合人类使用,以及我们该如何从人类的角度去理解 AI 系统,让它们以符合人类价值的方式服务于人类的目的。这就意味着,我会特别关注魏岑鲍姆的论证,并且援引当前维也纳正在进行的有关数字人文主义的讨论(Werthner, 2019)。

我将聚焦于 ChatGPT,仅把它视为一个知识资源。本文是多次实验—反思循环的成果,借助这些实验与反思,我评估了该系统的优缺点,并总结出一些讨论它们的方法。我开始意识到 ChatGPT 在回应看似无限多样的提示时那令人惊叹的自信,以及它用极为清晰的语言表达回应的能力,不管这些回应准确与否。

在下文中,我首先会回顾 ChatGPT 之前的技术发展,这有助于刚涉足该领域的人理解基于深度学习的 AI 语言模型的性质。其次,我将反思自己为探究系统“内部运作”、了解其功能和识别其极限所做的一些实验。基于这些实验,我会阐述自己对这类系统所提出的最大挑战的看法,即它提出了一种新的理解“真”——或者说废除“真”——的方式,我称之为通过相关性来理解的“真”。最后,我将指出人类为应对这一新挑战可以采用的思考和行动方式。

二、从 ELIZA 到 ChatGPT 的漫长道路

ELIZA 系统是首个展现出某种语言理解能力的程序,这种理解能力使人与机器的对话能够表现出类人行为。不过,“对话”这一术语具有误导性:虽然它通常表示人与人之间亲自进行或者通过通信媒体进行的自由口头或书面交流,但在这里它指的是一系列人的提示和系统的响应。

对于魏岑鲍姆而言,像 ELIZA 和 ChatGPT 之类的系统,只有那些不了解其内部运行机制的人才会觉得神奇。ELIZA 就是如此,它是一个独立的、基于规则的程序。其行为虽然由于逻辑规则间复杂的相互作用而难以预测,但却可以重复。在一段时期内,同一输入得到的结果是稳定的。专家能够以令人满意的方式解释其内部运行情况。

现代聊天机器人建立在神经网络和深度学习的基础之上,在训练阶段,它们会接触海量的文本语料库,吸收这些语料库,并将其作为回应的依据。我们无法从规则的角度解释它们如何针对给定输入生成结果。相反,这些结果是通过对文本模式和相关性进行统计分析得到的。并且,它们会基于用于训练的特定文本语料库和查询序列获得的反馈随时间演变。它们的机制不为人知,就连最有经验的观察者也无法预测其针对给定输入输出的结果。

将 ELIZA 和 ChatGPT 做比较,就如同比较黑森林里的多瑙河细流与塞尔维亚的浩瀚大河。它们之间的差异不能仅从多瑙河源头的角度来解释。相反,后者的浩瀚汹涌是多种地理上不相干的因素共同作用的结果。要理解多瑙河这个整体系统,就必须将中东欧的广大区域考虑进去。下一节将尝试对 ChatGPT 做同样的事。

(一) 围绕 AI 的技术

从依赖本地(非联网)计算机的硬件和操作系统的独立系统起步,从 ELIZA 发展到 ChatGPT 大约花了 57 年时间。要理解这一巨大的技术进步,我们就必须思考在这几十年间多个科学学科领域所取得的进展,包括理论基础的建立、技术成果以及这些学科之间的有效互动。

与计算机相关的语言研究起步很早。用计算领域的行话来讲,我们所说、所写的语言被称为自然语言。1960 年左右,计算机语言学这一研究自然语言的新领域应运而生。在它的推动下,计算系与语言学、文学和媒介研究领域开展了大量合作。索加德(Søgaard, 2022)认为,ChatGPT 并非科学贡献,而是工程壮举。在此语境下,ChatGPT 主要得益于计算机导向的语言研究学科的进步,其语言技能的确令人钦佩。

近几十年来,计算机技术取得了巨大进步。其操作模式虽与原来相同,但计算能力提高了好几个量级。计算能力的增长,使 AI 能从原本基于规则的模型转向神经元网络,而神经元网络正是 ChatGPT 的技术基础。

20 世纪 80 年代,人机互动在软硬件层面均得到研究,并取得了革命性发展。于是,便有了如今我们理所当然接受的互动模式和风格,现代聊天机器人的前身也随之产生。

网络技术的发展始于 20 世纪 60 年代。最初,人们为了军事目的将计算机关联起来。到了 20 世纪 70 年代,一场革命催生了新的范式,按照这一范式的理解,网络应通过计算机将人关联起来。在广泛的研究、技术创新和标准化的推动下,互联网以及新的线上工作与通信基础设施出现了。

在 1990 年左右,万维网向公众开放,知识产品经数字化后可为所有人所用,这从根本上改变了人类的知识文化。线上知识产品的全球发布也应运而生,随之出现了各种积累和分享人类知识的新形式。

不久之后,数据科学出现了。它作为计算的一个具体视角,聚焦于存储、组织和检索文本、数字和影像形式的大型数据库。在开发出搜索、分类、诠释和分析的自动方式后,人们产生了一种想法,即可以从数据分析中得出一种新形式的知识。

同时,统计建模与学习算法的发展也从概率角度对数据科学和数据分析产生了影响。其思路是找出并比较相关文本或影像中的模式,从而能够以极高准确率预测它们之间的关联。从本质上讲,支撑 ChatGPT 架构的正是概率方法。

这个清单或许并不完整,但仍旧令人印象深刻。七个学科,每个学科都有其复杂的研究议程、创新技术和广泛影响。我们必须将它们全部考虑在内,这样才能理解自 1966 年以来 AI 的巨大进步。

(二) AI 领域的范式转换

我们需要在这样的背景下,通过与多个计算学科进步的持续互动,理解 AI 领域自身的变化。必须记住,AI 是一个本质上跨学科的研究领域,它与认知科学密切交织在一起。

也许在 AI 领域,最重要的变化是从符号方法(基于明确的逻辑规则来思考思想和认知)向神经元网络方法转变,后者试图模拟大脑中神经元的组合与神经元之间的关联。值得注意的是,在符号阶段,AI 的所有断言都能从逻辑角度表达,而神经元方法是次符号的(sub-symbolic),它完全依赖数学属性和神经元与神经元突触的当前状态。

这一点尤其适用于机器学习。机器学习基于这样一种想法:AI 系统可以通过经验,即接触样本数据和反馈,超越制造者赋予的知识。

在基于规则的学习中,系统可通过应用所给规则来扩充知识库,并且存在形成新规则的机制。另一种可替代基于规则学习的是使用神经元网络的

自适应学习(adaptive learning)。自适应学习仅关注神经元和神经元突触的状态,此状态初始被设为初始值。随后,系统在训练阶段接触样本数据,并根据在数据中的发现来调整自身状态。在这种模式下,有监督学习(supervised learning)涉及人类训练者提供的反馈,而无监督学习(unsupervised learning)是让"算法只从未标记的数据学习模式"(Wikipedia)。自适应机器学习的质量取决于网络架构,包括神经元数量、排列以及从原始输入到学习输出所经过的层级数量。

(三) Transformer 架构和深度学习

在这里我得老实承认,我没资格写下面这部分内容。我查过一些原始文献,也乐意欣赏同事们的工作,但我不确定能否用自己的话来介绍他们正在做的事,即便我自认为是个聪明的观察者。

古德费洛等人(Goodfellow, Bengio & Courville et al., 2016)的作品对深度学习这一主题进行了全面探讨,建议感兴趣的读者去参看该文本。不过,为介绍这一概念,请允许我引用金为那本书所写书评开头的一段话:

> 深度学习是机器学习的一种形式,它能让计算机从经验中学习,从概念层级的角度理解世界。计算机通过收集经验知识,无须人类操作员正式阐明所需的全部知识。概念层级使计算机能够从较简单的概念构建出复杂概念以学会复杂概念;这些概念的图形表达会有多个层级的深度(……)事实证明,深度学习在许多软件学科中都很有用,包括计算机视觉、语音和音频处理、自然语言处理、机器人学、生物信息学和化学、视频游戏、搜索引擎、线上广告和金融。(Kim 2016: 351)

也就是说,负责深度学习的机制借鉴了多个数学学科(特别是统计学

和线性代数),以形成合适的学习算法,用来合成系统学习策略。

瓦斯瓦尼等人(Vaswani, Shazeer & Parmar et al., 2017: 5999)引入了 ChatGPT 所基于的 Transformer 架构。他们在导言中写道:“在这部作品中,我们提出了 Transformer,这是一种完全依靠注意力机制来绘制输入与输出之间全局依赖关系的模型架构。”

Transformer 架构包含两部分:一是将输入序列映射为语素内部向量表示的编码器,二是把该向量转换为输出序列符号的解码器。这两部分均具有六个相同的层级。为根据输入序列计算输出,Transformer 使用一种名为自注意的机制,此机制逐点检查向量,并将中间结果并入输入以进行进一步检查。对本文的讨论而言,有这样一个概念框架足够了。

从用户的视角出发,我们来做几点评论。若我在学习一个系统的操作模式时遇到困难,那么其他人又会遇到什么困难呢?获取 ChatGPT 的实操知识需要花费一些精力,将其传授给没有经验的用户更是极为费力,还可能引发误解和过高的期望。主要面临以下几个挑战:首先,系统的学习过程不透明,它依赖像学习算法和标量向量积之类的数学运算,这些从用户角度无法解释;其次,系统的回应完全取决于所选的文本语料库以及在训练阶段接触文本的顺序;再次,由于它是无监督学习,因此就连制造者或训练者也无法预测结果,也就是系统的知识库;最后,除了让它接触更多数据外,我们没有任何办法去了解或影响系统的知识库,并且这样做的后果也是无法预测的。

因此,仅训练阶段就引出了几个引人深思的严肃问题:应如何选择材料?如何将不受欢迎的态度排除在知识库外?如何在不同观点之间保持公正?

而且,还存在资源消耗的问题。将语素转化为巨大向量、在向量空间进行数学运算以及生成输出,这些都需要海量的处理能力和 GPU 时间。这是

Transformer 方法本身就存在的问题,并非特定系统才会产生的后果。所以,如果你正在使用 ChatGPT,请注意,从能耗的角度来说,你就像是在驾驶一辆 SUV。

三、和 ChatGPT 打交道的经验

本文本质上是对我在 2023 年 4—8 月使用 ChatGPT 的经验的反思。我提出了一些问题,涉及各种各样的话题。我通过让它写诗来探索其创造能力。我有一些意外发现,比如,不管什么原因,当问题超出其能力范围时,系统就会变得不稳定。系统的能力范围从未被明确说明过。当你要求系统说清楚时,系统会深表歉意,并最终感谢你提供宝贵反馈(这一反馈应对开发者有所帮助)。在陷入沟通僵局时,会突然有人在后台介入,重新引导系统。

(一) 关于引用 ChatGPT 的陈述

在这里,我会引用与系统实际对话的文字来使我的论证更具体。不过,读者应注意,按寻常方式引用 ChatGPT 是不可行的。因为引文没有稳定来源,与 ChatGPT 的互动实验不可重复。在 ChatGPT 中,系统的行为取决于询问的具体措辞以及聊天历史,并且会随时间而改变。为给我的回应提供语境,我得重现整个聊天过程。与人的访谈不同,我无法将对话当作证据记录下来。这样一来,我只能依靠读者的意愿了,希望读者相信我引用的文字来自与 ChatGPT 的实际对话。我不能用这些引文来证明我的论证,它们仅起说明我想法的作用。

为给我在本文开头引言中所用引文提供一个语境,我得说明我是如何促使系统做出这一惊人招供的。这一互动发生于 2024 年 4 月 3 日,在一次提出不同问题的聊天即将结束时:

CFL：我知道我什么都不知道。这对你来说有什么意义吗？

ChatGPT："我知道我什么都不知道"这句话常被认为出自古希腊哲学家苏格拉底。它体现了一种谦逊的哲学态度和智识上的谦虚，即承认人的知识有限、存在错误和不确定性。

作为 AI 语言模型，我不像人类那样拥有知识，也不具备自我反思或内省的能力。不过，承认我们的局限性，承认错误与不确定的可能性，这对 AI 技术的发展和使用仍具有现实意义。

AI 系统是强大的工具，能以多种方式帮助人类。但它们并非绝对正确，可能犯错或错误解读语境。对 AI 技术的用户和开发者而言，认识到这些局限，以谦逊和智识上的谦虚来使用 AI，在继续开发和改进这些强大工具的同时不断努力学习和进步，这很重要。

这个回应的结构很能说明问题。第一段与我的询问相关，作为答复已经足够。但某些因素（可能是"知道"这个关键词或者之前的聊天历史）触发了系统，导致它又添加了两段话，将论证延伸到了我未要求的程度。第二段提供了关于 AI 语言模型的人工自我反思，好像它在像一个自主的行动者那样行动。然而，第三段则要求人们关注 ChatGPT 涉及的人，建议人们在开发 AI 的过程中对自己的角色保持谦逊。

我最近再次复制了自己的提示，得到的回应却截然不同。所以，我无法保证，如果你也这么做系统会给出相同的回应。（也许你仍会好奇，想要尝试一下。）

（二）与机器对话

注意，在前面讨论的例子里，我将 ChatGPT 称为"你"。之后我不会再这么做了，因为我不觉得自己和机器有这样的关系。不过，ChatGPT 的程序

仍使其以第一人称回应。

在所谓的对话里,魏岑鲍姆的 ELIZA 本质上是基于已知关键词,按照规则转换用户输入的。ELIZA 只会偶尔以“我”的身份回应。即便这种在当时相对原始的输入—输出转换,也让很多用户产生一种感觉,仿佛系统像人一样,是一个值得与其分享秘密、被认为有自身动机和价值的对象。ChatGPT 也是把提示文本转换为回应,就更是如此了。不过,利用复杂的统计学方法,转换策略已经变得复杂得多了。

大多数 AI 开发者的方法是让自己的系统尽可能像人。所以,ChatGPT 大量使用第一人称短语,系统模仿人类礼貌形式时的情况特别令人震惊。(例如:“我为之前回应的不准确而道歉。你是对的,感谢你的耐心。”)不管我愿不愿意,都会被拉进看似对话的情景,我必须一直记住,得到的回应是人工生成的。

有一次,我问系统:“怎么会有人说人是机器呢?这属于范畴错误啊。”随后我得到了如下回应:

> 你说得对,人和机器分属不同范畴。人是生物有机体,机器是人造物,所以说人从字面意义上讲是机器属于范畴错误。不过,也可将人是机器这一想法理解为一种隐喻或者概念性的说法。

从严格意义上的历史来看,人只与他人对话。当然,在宗教、神话以及童话故事里,更高的存在也会用“我”来介绍自己。不过,在这种情况下,对话是无法进行的。如今,我们很多人选择机器这一隐喻,并且毫不犹豫地参与与“人造物”(就像 ChatGPT 在描述自己时所说的那样)的对话。

（三）探索不同的话题

在我孙子的帮助下，我开始与 ChatGPT 聊天玩。没想到，马上就得到了一个不尽如人意的回应。受我即将到来的暑假计划的启发，我用德语询问："北波兰（Nordpolen）"（波兰北部）7 月份的气候如何？系统却认真地回复了四段有关北极（Nordpol）的内容。我孙子把输入改为"北-波兰"，才得到了相关的回应。这个微小而明显的错误仅仅表明，系统的德语词汇有限，超出这个范围就会产生错误信息。（在后来的聊天中，ChatGPT 解释说，"北波兰"根本不是一个地理区域。）

其他错误和局限就没那么容易解释了，因为它们取决于用于训练的文本语料库以及提示的具体措辞。

最近几个月，我和 ChatGPT 聊了很多话题，这里列举部分例子，像北波兰作为一个地理区域、克里斯蒂安娜·弗洛伊德的简历、利用无人机发现小蠹虫、一战的历史、尼古丁和大麻的危害，还有"人工智障"等话题。

我最初感兴趣的是探究 ChatGPT 对不同话题各方面呈现中的系统知识范围、准确性和偏见。随着探究过程中更多问题的出现，我的研究也变得更加系统。

首先，我开始担忧回应的可信度：若事先不知答案，我能否辨别回应是否准确、相关？为何系统有时会产生错误信息？我能否分清可靠回应与任意回应之间的界限呢？

其次，在我纠正系统时，我开始对系统的反应感兴趣，即 ChatGPT 如何应对自己的错误。

再次，我想知道 ChatGPT 为何会忽视某些话题相关且必不可少的信息，又为何在同一话题的后续询问中自相矛盾。具体而言，我想知道我的反馈是否会对后续互动产生影响。

最后,我开始好奇,复杂的问题是否得到了全面涵盖。例如,关于内容的报告是如何结构化的,提及了哪些内容,忽视了哪些内容,是否存在文化偏见。

我也开始对从以下角度了解系统本身感兴趣:应用于人和机器的知识概念、ChatGPT 关于自身应用的自评估、系统在理解和领悟方面的能力和局限。以下是我的反思,我将用一些聊天样例加以说明。

(四)内容:准确还是任意?

2023 年 3 月初,我听闻 ChatGPT 给出了有关我的奇怪信息。受此触动,我就自己的生平事实展开了一系列的聊天交流。我得到的回应千差万别,在基本信息方面(像国籍、生日、出生地、科学贡献、雇主、所获奖项等)大多是错误的。按照 ChatGPT 的说法,我出生在德国的多个地方,还被视作包括形式方法和软件度量在内的多个领域的专家。例如,2023 年 5 月 7 日,我被告知,克里斯蒂安娜·弗洛伊德是一位出生于莱比锡且于 2021 年去世的德国计算机科学家。对此,我对该系统提出了质疑(以下文字由我翻译成英文):

CFL:又一次提到克里斯蒂安娜·弗洛伊德出生于路德维希港。ChatGPT 怎么会在同一话题的连续询问中自相矛盾呢?

ChatGPT:很抱歉我之前矛盾的答复让你感到困惑……不过,我会尽力对你的询问给出准确、相关的答复……

CFL:实际上,克里斯蒂安娜既不出生于路德维希港,也不出生于莱比锡,而是出生于维也纳。她不是德国人,而是奥地利人。她并没有死于 2021 年,就在当下,她正在给 ChatGPT 写几行字呢。

ChatGPT:对不起,我上一个答复中的信息有误。实际上,克里斯

蒂安娜·弗洛伊德于 1943 年出生在奥地利维也纳,她是奥地利公民……此外,我想澄清一下,克里斯蒂安娜·弗洛伊德仍然健在。我之前的答复不准确,很抱歉给您造成了困扰。

然而,这种自我纠正并未给系统留下长久的印象。5 月 8 日,我询问克里斯蒂安娜·弗洛伊德是否仍然健在:

ChatGPT:不幸的是,克里斯蒂安娜·弗洛伊德于 2021 年 6 月逝世,享年 77 岁……

CFL:你从哪得知克里斯蒂安娜·弗洛伊德已经去世的?

ChatGPT:……数家知名新闻渠道以及包括柏林工业大学在内的学术机构,广泛报道了克里斯蒂安娜·弗洛伊德于 2021 年 6 月 14 日去世的消息……

CFL:真不可思议,我是克里斯蒂安娜·弗洛伊德,我还活着。

ChatGPT:我为之前回应中的错误道歉……很高兴得知您还健在,弗洛伊德教授……也为之前回应可能造成的所有困惑或痛苦致歉……

为测试系统是否记住了我的聊天内容,并且能否从我的抗议中学习,2023 年 5 月 9 日,我再次发起了一次聊天:

CFL:和我说说克里斯蒂安娜·弗洛伊德的情况吧。

ChatGPT:克里斯蒂安娜·弗洛伊德(1943—2021)为德国计算机科学家,是软件工程领域的先驱。她出生于德国特劳恩施泰因……

路德维希港、莱比锡、特劳恩施泰因——可我坚持自己出生于维也纳。

显然,系统对与主题相关的诸多文本并不熟悉,所以无法建立可靠模式。它想确定一个出生地,但结果很随意:由于我被视作德国人,在它看来,德国的任何一座城市似乎都可以。

(五) 结构和重点

ChatGPT 的惊人特征之一,在于它能够将复杂话题的回应结构化。不管你询问的是利用无人机寻找小蠹虫、一战的进程,还是尼古丁和大麻的危害,该系统往往都会先生成一个开头段落,接着是一系列带有编号的段落,每个段落探讨一个子话题,最后是一个进行总结论证的结尾段落。如果让它生成报告,它甚至还会提供摘要。这种结构依赖于系统查找和命名合适方面或子话题的强大能力。

如果我是学校教师,要判断学生交的报告是否由 ChatGPT 撰写,我会很快学会识别这种结构。下面这个经过删减的例子应该能说明这个基本想法:

CFL:我对一战的总体描述感兴趣。

ChatGPT:一战……从 1914 年持续到 1918 年……涉及两大同盟:同盟国(主要包括法国、俄国、英国,后来美国等国家也加入其中)和协约国(德国、奥匈帝国、奥斯曼帝国和保加利亚)……

一战的主要方面如下:

1) 原因:……政治、经济和民族等多种因素相互交织。1914 年奥匈帝国斐迪南大公遇刺……

2) 堑壕战:西线发生了堑壕战……

3) 技术创新:……对新型毁灭性军事技术进行了有效利用……

4) 僵局与战斗:战争在西线迅速陷入僵局……带来了巨大的人员

伤亡并展现出战争的恐怖。

5）全球冲突：一战虽然主要集中在西线，但也是一场全球性冲突，战斗在多条战线展开，如东线（俄国与德/奥匈帝国之间）、意大利战线以及中东、非洲和亚洲的各个战区。

6）人员伤亡：……

7）帝国的终结：……

8）《凡尔赛条约》：1919 年，战争以《凡尔赛条约》的签订正式宣告结束……

9）遗产和影响：……

总之，一战是一场全球性冲突，对民族国家、社会和个人都产生了深远影响……

在提交提示之后几乎立刻就能得到这样一份概述，这确实令人印象深刻。我对这个话题也相当熟悉，可速度绝对比不上它。不过，值得研究的是，它是否经得起细查。

（六）语境和视角

一战的“总体”描述无法令我满意。我发现其中遗漏了重要参与方，并且对战争多个方面的呈现也不准确。所以，我想要从所有参与方的视角逐一进行阐述（“我对从［……］的观点来看一战是什么样子的感兴趣”）。

当我问到哈布斯堡帝国时，我着实被震惊了，因为 ChatGPT 没有提及斐迪南大公遇刺一事。在常规的深表歉意之后，我得到了关于这一事件重要性的详尽解释。不过，在 ChatGPT 看来，这一事件与奥地利的历史似乎并无内在纠葛。

这引发了一些更广泛的疑虑：我不太了解奥斯曼帝国以及一战涉及的

其他一些国家的历史。要确保从各个角度涵盖所有相关子话题,似乎是不可能的。

我的结论是,ChatGPT 在训练过程中以英美视角为主,其语境完全依赖该视角。

(七) 本体论层面上的和文化的偏见

为了测试文化偏见,我问了几个历史方面的话题。历史往往是从一个国家或者一种文化的视角去讲述的。当我们从一个国家到另一个国家时,我们眼中好的东西可能就变成坏的了,过去具有现实意义的东西可能也不再有现实意义了。

比如说,从我的中欧视角出发,思考一下关于一战的“总体”描述会是怎样的情况。在引言部分,主要参与方之一的意大利未被提及。第 2 点和第 4 点存在片面性,忽视了其他战线上(如奥匈帝国和意大利之间的伊松佐河沿线)的堑壕战。战争导致各地都有大量人员伤亡。第 5 点的偏见也很明显:“主要集中在西线。”东线和意大利战线本就是“全球冲突”的一部分。然而,意大利北部、巴尔干地区和喀尔巴阡山脉等地——这些都是战斗发生的地方——与法国和比利时一样,都属于欧洲。第 8 点仅提及《凡尔赛条约》这一与德国的协定。然而,巴黎和会签署了更多影响不同国家的条约,包括《圣日耳曼昂莱条约》(针对奥地利)、《特里亚农条约》(针对匈牙利)、《塞纳河畔讷伊条约》(针对保加利亚)以及《色佛尔条约》(针对奥斯曼帝国)。

因此,ChatGPT 对一战的描述明显存在各种偏见:在选择实体、话题、事实与事件时,它会在呈现中进行包含或排除操作,对挑选出的一些子话题或方面赋予不同的重要性,并且给事实或事件赋予正面或负面价值。这种偏见是由其接触的以美国为中心的文本语料库所导致的。

任何人在任何时间、任何地点所写的任何文本都存在偏见。然而,基本的差异在于,由人创作的文本是有作者的,所以我们能够将偏见追溯到影响该文本创作的社会、历史或者智识条件上。与之形成对照的是,AI 语言模型生成的回应看似普遍且客观,但其中隐含的偏见无处不在,只是未被承认罢了。

(八) 检查结果的正确性

虽然大多数时候我只问自己知道答案的问题,但很快我就意识到,不能未经仔细核查就接受 ChatGPT 的回应。即便使用的是优质且资源丰富的工具,我也不能轻信它的结果。于是,我转而去使用我的浏览器、搜索引擎和 Wikipedia。

为了好玩,我又回到“北波兰”这个话题,在浏览器搜索框中简单输入“Nordpolen, Klima, Juli”(北波兰,气候,7 月)这几个词。0.38 秒后,网页上就出现了一个信息丰富的答案:7 月是去波兰北部旅游的好时节,那里属于海洋性气候,平均气温 25℃。这个答案下方是一个列表,列出了提供更多详情文章的链接。整个过程简单流畅,非常自然。毫无疑问,这个语义网搜索也涉及大量 AI 技术。

用无人机寻找小蠹虫,同样,简单搜索也得到了满意结果。关于尼古丁和大麻的危害,我甚至被警告,应该要么拿大麻与烟草相比,要么拿尼古丁与四氢大麻酚相比。

同时,我从 Wikipedia 获取的信息要比从 ChatGPT 获取的信息丰富得多,而且 Wikipedia 还能帮我关联到相关信息的原始出处。

(九) ChatGPT 作为知识资源

我开始把 ChatGPT 生成的回应当作死文本(dead texts)来思考。就当

前讨论而言,当一个文本嵌入人类话语和活动中时,我将称它为活的,反之则为死的。这种对比对于 ChatGPT 的回应和其他知识资源来说具有现实意义。

例如,Wikipedia 的文章也属于知识资源。其旨在呈现知识,具体方式为澄清概念、给出相关历史背景、总结关键结果,最终通过指出作者、链接其他文章、参考其他文献(包括一手文献)来丰富人类话语。所以,它们是鲜活的。

ChatGPT 坚称,作为一个 AI 语言模型,“[它]不具备人类那样的知识”。不过,值得探究的是,ChatGPT 以何种方式“占有知识”。当被视为知识资源时,ChatGPT 的回应难以令人满意。通过系统的权威,你得到你所得的,它就是它本身的样子,它可能准确也可能不准确,接不接受取决于你。ChatGPT 回应中所包含的知识就像肉末(ground meat):你无法分辨其来源,而且获取过程也不透明。这到底是谁的知识呢?因此,我提议将其称为“知识末”(ground knowledge)。

知识末不能用于论证,因为它缺乏来源且存在短命问题——ChatGPT 下次回应相同问题时可能会给出不同答案。所以,若要在现实语境中使用 ChatGPT,就必须借助学习共同体中的评论与诠释将其回应变“活”,这对 ChatGPT 而言比对其他知识资源更为必要。ChatGPT 的回应至多只能作为讨论的合适起点,后续仍需人类对其进行修正。

四、通过相关性来理解的“真”?

在进行实验与反思时,我开始发觉 ChatGPT 的关键问题在于其对待“真”的方式,“真”的理念并不适用于 ChatGPT 的内部运作。

与旨在让陈述为真不同,系统关注的是一个陈述准确的概率。准确性

是避开“真”的一种方式。说一个陈述有 90%的概率准确是合理的。系统的设计基于这样一个假设:如果该概率足够高,甚至高于一个人回答同一个问题准确的概率,那么人们就会愿意放弃“真”,而接受“高概率”或者“近乎于确定”来作为替代。

(一) 我们可以不要“真”吗?

在英语里,“真”关联着一个极为广阔的语义场,其含义包括:对他人有信心层面的“信任”、真诚可靠意义上的“真”、个人态度或制度政策意义上的“真实性”。有一种说法称,“‘真’是自由和民主最好的保障,也是真正的基础”。我们的日常生活依赖于一个条件,即能够依靠准确的信息。在法律诉讼中,若不坚持“真”,法律和秩序就会崩溃。没有交易各方之间的信任,任何商业交易都无法进行。即便我们身处一个“假”的时代,“真”和真实性仍然是支撑人们在社会中共同生活的精神支柱。

形形色色的哲学流派提出了各不相同的关于“真”的理论。也许,并不存在绝对意义上的“真”,但对“真”的追求是人类生活的一大支柱。

(二) 没有意义的相关性

ChatGPT 通过匹配输入文本中的模式与从训练文本语料库中已知的模式,来建立“真”的概率替代(或者更确切地说,建立“真”)。最佳输出是根据文本相关性预测出来的,即根据变量间的相互依赖(此处是将输入文本中的语速与从训练数据得出的语言模型做比较)预测出来的。

简单来说,将一个词或语素作为下一项纳入输出文本的依据,是该词在此处出现的概率,这个概率由系统通过接触训练文本语料库所得到的所谓“经验”确定。

（三）没有作者身份，没有论证

对“真”的追求与作者身份相关，且需通过论证达成。论证源于个人在世界中的体验，需要经验证据与逻辑推理。我们要对自己陈述的真实性负责，没有证明就不存在“真”，我们必须互相说服。

论证就如同跳舞，舞伴会轮流表明自己的立场，并且邀请他人加入。在舞蹈里，一致与冲突都是能引出下一步的有效舞步。

就像我所展示的，你无法通过引用 ChatGPT 的陈述来论证，因为 ChatGPT 的陈述没有作者，不可重复且容易变化。在 ChatGPT 那里，所有原创作者的知识都被混合在一起，绞成了肉末。

而且，试图与 ChatGPT 争论只会令人深感挫败。若你表示反对，它便会给出看似礼貌的回应，接着把你输入的内容重新措辞后再重复一遍。你可以不断重复这一过程，但互动风格不会改变。这就像是一场堕落的舞蹈。

五、“真理解”的缺乏

ChatGPT 使用“真理解”这一术语来描述自身的能力与局限。当时，我在探寻各种方法，以便从它的理解和知识层面更清晰地认识 ChatGPT。我想要弄清楚，我们通常用于人类的概念能如何应用于 ChatGPT。当我直接提出这个问题时，它在解释自身内部运作时给出的答案清晰又准确，这让我印象深刻。事实证明，询问关于理解的问题格外有用，2023 年 8 月 28 日的一次聊天就得出了这样一个富有启发意义的自我评估：

CFL：ChatGPT 有怎样的理解？

ChatGPT：ChatGPT 的理解基于从训练文本中学到的模式。它并

非像人类那样真正理解,而是能够识别并模仿训练数据中的语言、信息和语境模式。其理解机制如下:

1)语境理解:ChatGPT 能够生成在语境上看似相关的文本,原因在于它能识别语言中的特定模式。它会考虑之前的对话历史,生成在该语境下连贯的回应,以此来理解对话语境。

2)文本的相似性:ChatGPT 能够识别并复制在训练期间遇到的相似短语、想法和概念。如果用户输入与它所见过的内容相似,它就会给出看似相关的回应。

3)事实信息:ChatGPT 能够提供在其训练数据中出现过的事实信息。然而,它无法独立验证事实,也没有可借鉴的现实世界经验。

4)有限推理:ChatGPT 能够基于所学模式执行简单的推理任务。例如,它能够遵循基本逻辑,回答涉及直接演绎或推论的问题。

5)语言模式:ChatGPT 擅长生成符合语法规则和语言模式的文本。它能生成听起来流畅连贯的文本,即便内容本身可能未被深入理解。

不过,ChatGPT 的理解存在一些局限:

1)缺乏真正理解:ChatGPT 没有真正的理解和意识,无法把握抽象概念或情感。

2)缺乏常识与分辨语境微妙差别的能力:它难以理解微妙的语境、反讽、讽刺或隐喻,有时提供的回应看似合适,却缺乏深度理解。

3)存在错误信息与偏见:若训练数据中存在某些模式,ChatGPT 会无意生成不正确或有偏见的信息,它不具备独立验证信息的能力。

4)没有个人经验:它没有个人经验、情感或意识,其回应基于训练数据中的模式生成。

总的来说,ChatGPT 的理解仅局限于识别和复制其训练数据中的

语言和信息模式,在看待它的回应时,应牢记这些局限。

这个描述阐释了系统以合法方式成功构建文本模式的策略。输出需与输入的询问“相关”,即应收集可从文本推断出的、与用户心中所想相关的方面。并且,得出的文本要“连贯”,也就是文本间应相互关联而不自相矛盾。这两个特征(“相关”和“连贯”)虽颇为可取,但不一定与现实世界相关。

(一) 无理解的理解

正如之前所讨论的那样,“ChatGPT……并非像人类那样真正理解概念,不过它能够识别并模仿模式……”

ChatGPT 表明,缺乏“真”理解或“真正的”理解是其第一个局限。在尝试获取一些相关背景知识时,我找到了索加德(Søgaard, 2022)的一篇文章,并且了解到人们就具有 Transformer 架构的 AI 语言模型是否理解语言展开了激烈的理论争论。这是一个关于理解与语义之间关联的问题。

ChatGPT 得出的理解是一种基于推论语义(inferential semantics)的理解:它能够识别并模仿语言中的模式。然而,ChatGPT 本身并不存在任何推论语义,因为其陈述并非基于现实世界的经验。

按照一个思想流派的观点,缺乏推论语义的 ChatGPT 确实无法达成真正的理解。不过,也有人提出了一个有趣的反驳。神经科学假设,我们在大脑的神经元和神经元突触中构建我们的世界地图,而 ChatGPT 依赖的是由语言模式构建的地图。为什么由词汇使用频率形成的语言模式和我们大脑中由世界经验频率形成的模式不能是同构(isomorphic)的呢?这对哲学家来说无疑是一个极具吸引力的问题。

对于 ChatGPT 用户而言,这个问题可简化为:机器基于模式的理解,能

在多大程度上接近人类基于经验的“真”理解？在何种情况下，机器的理解可被人类知识取代，又在何种情况下能取代人类知识？另一方面，我们需要学习如何可靠地感知并应对二者的差异。说到底，唯有系统的经验能够提供指导。

（二）人为错误和人工智障

在寻找表达 ChatGPT 不时提供的骇人错误信息带给我的挫败感的方式时，我得知“人工智障”已经是一个公认的术语了。塔斯（Tas，2020）认为：

> 它被用作人工智能的反义词。不过，它有两层相反的含义：其一，机器学习算法在数据学习过程中犯了愚蠢的错误；其二，人工智能被降低了智能，从而会犯错，看起来更像人类。

在与 ChatGPT 打交道的过程中，我常遇到第一种情况，却没意识到第二种情况。这让我好奇，系统是否会在我没注意时降智。不管怎样，2023 年 5 月 9 日，当我让 ChatGPT 写一首名为《人工智障》的诗时，我还是被逗乐了：

> 在智能主宰的 AI 世界里，
> 有一个有趣却又令人苦恼的现象，
> 叫作“人工智障”，这是一种令人困惑的情况，
> 如此聪明的机器，却可能把事情搞砸。
>
> 它们被编写出来是用于学习、适应和演化的，

但有时,它们的算法不太搞得定,
不能应对人类思想与理性的复杂之处,
而且无论何时,它们都会犯错。

它们会错误地诠释一个词或一个短语,
或无法理解我们的表达,
我们的需要和需求最微妙的方式,
并最终做出让我们悲伤的事情。

当然,这不是它们的错,毕竟它们只是机器。
但退一步说,这还是会让人失望:
当我们依靠它们完成重要任务时,
它们却失败了,只能让我们收拾烂摊子。

所以,在我们前进的时候,让我们记住,
人工智能,虽然高级,
却依然倾向于犯错,就像你和我一样,
没关系,这只是我们现实的一部分。

迷人的是,这首诗叙事连贯、风格一致且节奏流畅。不过,我也注意到诗中存在一些将人和机器混淆的情况:这首由机器生成的诗用“我们”指代人,用“它们”指代机器。

我也反对“人工智能,虽然高级,却依然倾向于犯错,就像你和我一样”这种看法。人和机器所犯的错误是不一样的。虽然二者产生的结果相同(都是错误信息),但其起源、语境和意义却大不相同。

人为的错误需要放在我们的历史与当前情境的语境下理解。一般而言,这些错误源于我们做了想做却没做好的事。要将错误视为学习事件(learning events):我们能从自身错误中学习。在学习共同体里,错误可被当作让所有人获得更深入理解的机会加以利用。正如皮亚特·海恩(P. Hein)以极为精妙的方式所指出的那样:

> 通往智慧的道路——
> 哦,这听起来既朴素又简单:
> 犯错,犯错,再犯错,
> 但要错得少些,少些,再少些。(Hein, 2023)

一个知道我在柏林工作过的人也许会以为我出生在柏林。这也说得过去。可要是说我出生在路德维希港、莱比锡或者特劳恩施泰因呢?我非常确信,在用来训练 ChatGPT 的文本语料库中,没人写的档案会包含这种关于我出生地的惊人错误。

ChatGPT 被编写出来就会有这样的表现,它的行为是人类有意为之的结果,而那些意料之外的行为通常是人为错误间接导致的。不过,ChatGPT 经常给我们提供的错误信息并非人为错误造成的,而是基于模式的机制产生的。人工智能可能必然会导致人工智障(就像诗中所说:“没关系,这只是我们现实的一部分”),但我们必须意识到这一点。这些影响需要人类去检测、纠正、减少或者解决。

(三)如果系统不知道会怎样?

显然,ChatGPT 的局限性将开发者推到了台前。他们清楚系统的限度,也应该遵循一种一致的原则来减少“人工智障”现象。

我与 ChatGPT 打交道的经历让我欣赏人类的基本能力,包括:其一,承认我们有所不知;其二,将新见解纳入考量;其三,据此改变我们的想法。我希望,作为让 AI 系统更像人的下一步举措,这些非凡的能力能够在 AI 系统中得到模拟。

仅有一次,ChatGPT 主动承认自己知识匮乏。我问了些有关埃塞俄比亚的问题,这些问题明显偏离系统主流文化视角,然后被告知:“我找不到任何具体信息。”

与此同时,当我再次向系统提及北波兰时,终于得到这样一个解释:“可能会产生这种混淆,是因为您正在使用的 AI 语言模型是用大量德语文本数据集训练的,这些文本中可能没包含太多‘北波兰’这个词的实例。”没错,确实如此。可为什么不马上承认呢?

最后,当我提及反复出现的、有关我生活各方面的错误信息让我感到失望时,它坦言:“看来我对这个具体话题的知识有限且不完整。”不幸的是,我并不奢望这个坦白会让 ChatGPT 产生新的“见解”或“改变想法”。

六、对 AI 系统的负责任的使用

(一) 选择合适的工具

传统工具的目的可能一目了然,也可能容易学习,但先进的 AI 技术并非如此。OpenAI 的 PR 将 ChatGPT 与各种各样的目的联系起来,不过只要尝试一下,就会发现它的“有用性”存在问题。

在计算方面,我们在工具的选择上有很多。我通常会选择最简单且资源消耗最少的可用工具。例如,我已经展示过,Wikipedia 作为一个知识资源,有时能比 ChatGPT 更好地满足我的需求。要记住,我把 ChatGPT 比作

一辆 SUV：它能提供舒适和便利，但从存储和计算时间的角度来看，也会消耗大量资源。

ChatGPT 提供的舒适和便利体现在其统一的界面上，该界面让我们能够轻松地使用自然语言处理很多不同任务。不可否认，这个系统的语言技能相当出色，但在简单的询问中，没有这些技能也可以。

这个系统的长处和创新点在于其生成新的文本工件（如报告、论文、总结、诗歌、程序代码、翻译）的能力。保留 ChatGPT 用于这些更高级别的用途是值得的。正如我在 2022 年的文章中所讨论的，我认为这类工件属于“知识工件”（knowledge artifacts）：它们本身并不总是有趣，但能够用于发展先进的学习文化。

（二）系统的设计与评价

在日常生活里，我们使用许多技术时，并不了解其内部运作原理。不过，我们必须确保能可靠地理解这些技术在使用中产生的影响。这就需要我们借助 ChatGPT，从自身经验中学习，去了解它的能力、极限和缺陷。

讨论一般性 AI 的优点是存在问题的，因为基于 AI 的系统种类极为多样。问题不在于是否使用 AI，而在于如何针对人类的使用来设计基于 AI 的系统。

我们最基本的要求是系统可信、可靠、功能清晰且与人类目的关系明确。将这些标准应用于不同系统时，我们必须在讨论中将其具体化。显然，当前的 ChatGPT 并不符合这些标准，而这一点也应为修正、重新设计尤其是后续系统的设计提供指导方针。

关于设计的讨论是基于价值的，如为人类用户赋能或者维护安全。正如我们从其他部门所了解到的那样，在设计之初就必须明确并考虑人类价值。比如，不能等到建筑物建成之后才去考虑确保其安全。

虽然对于 AI 而言,具体的以人为中心的设计原则还有待提出,但我们可以从 IT 部门的其他技术中借鉴相关原则。以下是一些非常通用的原则:

小即是美:牢记舒马赫(Schumacher, 1973)的经典格言,这也是“像人很重要一样”设计技术的一条通用准则。它倡导我们以小型、目的明确、在最低限度消耗资源的同时松散耦合的系统为目标。

明确目的:我使用 ChatGPT 后得出结论,制造者野心过大会损害系统。谁会需要一个既能翻译语言、回答高深研究问题,又能处理和生成图像与文本、编写代码、写诗的工具呢?

限制范围:为什么聊天机器人要以解答世间万物为目标呢?为何不围绕一个话题领域打造小型聊天机器人,用相关文本训练它们,标明文本作者,并拒绝回答范围之外的问题呢?

确保可追溯性:目前,ChatGPT 的结论缺乏透明度。然而,对于严肃用途而言,可追溯性是一项强制性要求。我发现,系统能够对结论的得出过程予以解释,此功能必须明确地对所有人开放且便于使用。

启用纠错:用户无法纠正错误,这是绝不能被接受的。我明白,这不符合深度学习方法,深度学习的必然结果是系统只能通过接触更多文本来学习,即便如此,结果也无法得到保证;尽管这样,我们仍必须找到一种解决方案。

(三) 走向一种内嵌 ChatGPT 的学习文化

无疑,我们有充分的理由使用 ChatGPT 及类似系统。并且,我们会改进该系统,克服其当下存在的一些缺陷。

先进的 AI 系统促使我们围绕其发展一种更精致的学习文化,这是一种挑战。

也许,那些试图作弊或者只看重好成绩的学童会满足于 ChatGPT 生成的文本本身。我之前将 ChatGPT 的输出描述为“死文本”,我们必须让这种输出“活”起来:阅读、讨论、批评、检查然后修正它。如果我们想要在学习共同体(无论是学校、公司、媒体、政府还是卫生部门)中真正从 ChatGPT 获益,就必须认识到,AI 工具的主要价值不在于生成“死文本”,而在于生成能够促使人类产生更高水平学习过程的知识产物。只有这样,系统才能取得成果,实现质的增长。

我想针对在学习共同体中采用 ChatGPT 或类似系统的过程提出几条准则:

始终坚持真实性:不应满足于肤浅、误导性或错误的回应。要追问系统,向其要求更多信息,找出错误信息与自相矛盾之处,拒绝接受“假”的内容,并探讨应对“假”的方法。

努力提高人的能力:AI 系统是人的工具,并非目的本身。不要问它们能做什么,而要问我们能用它们做什么。要为 AI 系统的现实意义使用创造场景,创造学习使用它们的机会,在共同体内构建讨论、交流和诠释的空间,尊重、维护并提升人的能力。

强化责任结果:也许最为重要的是,将 AI 系统嵌入人类决策过程时需谨慎。例如,若患者在与医生见面之前要先与聊天机器人交流,那么医生应如何对诊断和治疗负责?我们还可将此问题延伸至医院责任链条及其之外的其他实例,以确保人对 AI 造成的结果负责。

最后,要永远记住,我们并非受 AI 摆布,相反,我们能够让 AI 服务于人类的使用需求。正如约瑟夫·魏岑鲍姆所说的那样,不要迷失自我。

参考文献

Floyd, C. 1986, "The Responsible Use of Computers: Where Do We Draw the Line?" *CPSR Newsletter* 3(2).

Goodfellow, I., Y. Bengio & A. Courville 2016, *Deep Learning*, Cambridge, MA.: MIT Press.

Hein, P. 2023, "The Road to Wisdom." https://en.wikipedia.org/wiki/Grook.

Kim, K. 2016, "Review of Goodfellow, I., Bengio, Y. and Courville, A.: Deep Learning." *Healthcare Information Research* 22(4).

Schumacher, E. 1973, *Small Is Beautiful: Economics as if People Mattered*, London: Blond & Briggs.

Søgaard, A. 2022, "Understanding Models Understanding Language." *Synthese* 200.

Tas, S. 2020, "How to Limit Artificial Stupidity." https://towardsdatascience.com/how-to-limit-artificial-stupidity-a4635a7967bc.

Vaswani, A., N. Shazeer & Parmar et al. 2017, "Attention Is All You Need." Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, New York, USA.

Weizenbaum, J. 1966, "ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine." *Communications of the ACM* 9(1).

Werthner, H. 2019, "Vienna Manifesto on Digital Humanism." https://caiml.dbai.tuwien.ac.at/dighum/dighum-manifesto.

数字平台帝国化的制度演进*

——评莱顿维塔的《云端帝国：数字平台如何重塑经济与世界》

牛一帆**

摘要：数字平台越来越表现出对内权力滥用、对外肆意扩张的帝国化倾向，成为诸多社会经济乱象的根源。基于在牛津互联网研究所的多年积累，威利·莱顿维塔的《云端帝国：数字平台如何重塑经济与世界》一书对数字平台帝国化的历史过程进行了深入探讨。本文首先分析了莱顿维塔的“云帝国”概念及其采用的制度历史分析视野，解释其中隐含的嵌套市场模型假设和制度变迁的异步性。在此基础上，呈现了数字平台通过建构电子商务市场制度而崛起的历史过程，以及面临市场利益相关者反抗所做出的制度调整。本文还将莱顿维塔的云帝国理论与技术封建主义、数字社会主义思想进行了对比，以说明该书在批判平台资本主义与提出替代性方案上的理论贡献。该书作为一本严谨、科学的学术专著，兼具可读性，但也因缺乏一个统领性的分析框架而存在一些不足。

关键词：数字平台　云帝国　制度历史分析　数字宪治主义

一、引言：作为云帝国的数字平台

当平台取代传统网页成为数据聚合与市场活动的主要场所时，平台资

* 本文系中国社会科学院学科建设“登峰战略”资助计划“科技与数字社会研究”（项目批准号：DF2023XXJC03）的阶段性研究成果。

** 牛一帆，中国社会科学院大学社会与民族学院、中国社会科学院大学数字中国研究中心。

本主义的进程便开启了。平台凭借其对“数据”这一生产要素的激活、提取和利用的技术优势,成为新的经济增长点,并迅速扩展到整个经济体系(斯尔尼塞克,2018)。在自身资本积累和扩张的同时(杜巧玲,2024;郑吉伟、初传凯,2022),平台也造成了新的经济和政治垄断(涂良川,2022),成为扰乱市场秩序(张雅俊,2023;谢尧雯,2021)、侵犯用户权益(蔡跃洲、王麒植、钟洲,2024;张谦、李冰晶,2021)、剥削数字劳工(马俊峰、赵海蕴,2024;Dommett, 2021; Rani & Furrer, 2021; Lin, 2021)等社会乱象的根源。

牛津大学互联网研究所威利·莱顿维塔(V. Lehdonvirta)教授在其新著《云端帝国:数字平台如何重塑经济与世界》(以下简称《云端帝国》)中,将数字平台类比为数字政府或虚拟国家(莱顿维塔,2024:IV),用以说明平台在当今资本主义发展和调节社会生活方面的重要作用。这本书自面世以来引发了学术热议。伦敦政治经济学院(LSE)第一时间发布了书评(Muldoon, 2022);英国期刊《社会经济学评论》(*Socio-Economic Review*)也专门为其开设了一个专题,邀请了 5 位知名学者以及莱顿维塔本人进一步澄清和深化书中的理论观点(Ergen, Antonucci & Arslan et al., 2024)。《云端帝国》作为一部较早讨论数字平台帝国化的专著,为理解平台权力的形成以及如何建立数字民主提供了创新性见解。

《云端帝国》的核心问题是数字平台如何走向帝国化:互联网并没有像最初承诺的那样,将人们从中心化的统治体系中解放出来,那么它是如何走向失败的?为什么硅谷的技术专家最终以数字形式重新创建了他们试图淘汰的中心化统治?莱顿维塔以制度历史分析的方法,讨论了数字平台崛起中技术与社会的张力:“我不会直接探讨技术是如何破坏现有的社会秩序,而是探讨它们如何构建一套完全不同的秩序。”(莱顿维塔,2024:XIII)他基于多年的研究和资料收集,每一章节都基于典型详尽、巧妙浓缩的案例研究,追溯了数字平台帝国的诞生和斗争史,并深入探讨了如何重新夺取数字

平台的统治权。

本文首先梳理和分析了莱顿维塔的“云帝国”概念及其制度分析的历史视野,特别是其划分的三种制度类型及其内在逻辑。其次,从数字平台对电子商务领域的制度供给,说明了云帝国崛起的历史过程,以及平台帝国因市场利益相关者所做出的制度权衡过程。最后,将云帝国理论放在平台资本主义批判的理论脉络中,将其与技术封建主义和数字社会主义进行比较,以分析莱顿维塔云帝国理论的学术价值与理论贡献。

二、云帝国的概念及制度分析的历史视野

(一) 作为云帝国的数字平台

将数字平台类比为国家,首先源于其巨大的用户规模和经济体量。有数据显示,脸书(Facebook)拥有超过 15 亿用户,其管理的“人口”比世界第一人口大国还多;谷歌处理了美国 64%的在线搜索和欧洲 90%的在线搜索,而阿里巴巴每年处理的交易额超过 1 万亿元人民币(1620 亿美元),占中国所有商业出货量的 70%(Gawer,2022)。更为重要的是,数字平台还拥有与之体量相匹配的巨大权力,数字平台越来越多地制定着我们生活中必须遵守的规则(莱顿维塔,2024:III)。

库尔德里等人(Couldry & Mejias, 2019: 38)更早地使用了“云帝国”(cloud empire)这一术语,用“云帝国”的概念来说明各类数字平台是如何通过数据对大众的社会生活进行持续、有组织的殖民的。过去的云帝国思想遵循平台帝国主义(Jin, 2015;严运楼,2020;温旭,2024)的理论脉络,虽然对于理解数字平台作为生产全球性不平等的技术根源具有启发性,但在这一理论中数字平台沦为发达国家殖民的技术工具,忽视了数字平台作为

经济主体的性质和能动性。莱顿维塔提出“云帝国”的概念,是想从内部管控和对外扩张两个维度批判平台权力。

在统治维度上,数字平台作为市场基础设施的提供者和服务者,享有管理内部市场、保证组织运行的私权力(private power)。有学者将数字平台管理 B 端和 C 端的权力场域称为内部治理场域(吕鹏、周旅军、范晓光,2022),数字平台在进行内部治理的过程中表现出管控属性(邵占鹏,2024)。莱顿维塔(2024:IV)认为,数字平台在内部治理的过程中表现出专制、集权的倾向:“这些虚拟‘国家’并非事事皆好……他们之中不乏有人滥用权力。”

在扩张维度上,数字平台凭借在数据和算法权力方面的优势,将曾经一部分属于政府的权力内部化为自己的权力(王志鹏、张祥建、涂景一,2016),造成了国家层面的权力流散(王沛楠、彭修彬,2023)。莱顿维塔认为,这是由于数字平台成功解决了许多国家面临的难题。一方面,数字平台替代政府成为新的信息处理中心,它所具有的技术优势在数据收集、统计计算、算法决策等方面的职能超越政府;另一方面,更重要的原因是,数字平台创建了跨国的制度基础设施,促进了更高效的全球贸易。数字平台的管辖权不被领土所束缚,任何在平台上注册的人——无论来自哪个国家——都可以通过平台实现跨国贸易。从这一意义上,数字平台是“没有地产的国家,是云中的帝国”(莱顿维塔,2024:256)。

(二) 制度变迁视野下的电子商务领域

“制度”是理解云帝国的核心概念。莱顿维塔(2024:V)开篇就对“制度”进行了明确而清晰的界定:“就本书所谓的制度……是指法律、法规、传统、行为标准以及社会中其他限定人们互动的‘游戏规则’。”莱顿维塔在制度经济学的脉络下,认为制度是解决市场交易中各种协调问题的一种方案。

从出于解决市场秩序和交易协调问题的理论层面讲,民族国家建立税收制度、推动密码学和计算机技术的发展,与数字平台建立数据库、发明区块链和进行算法决策在本质上是一样的。因此,数字平台与民族国家在建立制度和推动市场形成上具有历史似然性,这是莱顿维塔将平台与国家类比的合法性基础。

莱顿维塔区分了三类与数字平台内在相关的制度类型:第一类是经济制度,是指由数字平台制定的互联网市场的游戏规则;第二类是政治制度,是指改变这些市场规则并追究平台责任的机制;第三类是社会制度,是指为了保护和帮助互联网市场中人民所建立起的制度。从这三类制度的定义来看,莱顿维塔首先假定了一个嵌套的互联网市场模型,并将数字平台的角色定位于其中。首先是被数字平台管理的内部市场,数字平台制定着内部市场的所有规则,这是经济制度的维度;其次,这一内部市场又嵌入在更加广阔的电子商务领域之中,这一领域是该数字平台同其他平台相互竞争的细分行业市场,是充满着用户、商家、数字劳工等利益相关主体的权力场域,这是政治制度的维度;最后,作为内部市场统治者的数字平台和作为外部市场参与者的数字平台拥有着保护和保障用户基本生存权利的义务,这是社会制度的维度。

莱顿维塔采用的历史分析视角,在制度变迁的视域下展开了对云帝国崛起的分析。根据制度经济学的观点,制度变迁是指制度的替代、转化与交易过程(罗必良,2005),制度并不是一经生产就既定的,而是被组织和市场利益相关者相互作用而逐渐修改、演变的(诺斯,1995)。莱顿维塔一方面拒斥制度变迁的进化论,这一理论假设制度是朝向有效或高效方向发展的过程。他认为制度总是面临着特定历史情形中的交易协调问题,其有效性视具体解决的问题而定。另一方面,莱顿维塔也批判性地吸收了制度变迁的冲突论,他认同该理论所假设的制度总是在各种群体的权力斗争中被形

塑的,但并不认为制度总是既得利益者最大化自己份额的结果;相反,平台会在效率与公平之间做出权衡,动态地改变制度的规则。

在制度变迁的视域下,莱顿维塔将这三类制度置入了动态的、纵向的逻辑体系中。数字平台并非一开始就是一个“帝国”,它经历了一个历史过程。经济制度的分析发生在数字平台尚未崛起的互联网早期,是数字平台确立自己在互联网市场生态位、获得数字权力的历史时期;在数字平台崛起为一个中央集权的帝国之后,关于如何约束平台和改变现有规则的政治制度成为一个亟待解决的问题;随后,到了引起社会广泛关注和监督的平台转型期,数字平台如何建立电子市场的社会制度、履行自己作为市场公共主体的责任,是一个面向未来的倡议。这三类制度的发展并不是同步的,制度变迁的异步性造成了如今数字平台的制度堕距:数字平台在经济制度上处于 20 世纪 50 年代的计划与市场之间,在政治制度上处于黑暗的中世纪,而在社会制度上则仍处于未开化时期。

三、数字平台帝国化的制度溯因与权力变迁

(一) 权力起源:数字平台的市场制度建构

莱顿维塔通过对电子商务市场制度变迁的历史追溯,揭示了数字平台与国家在制度转型上的“历史似然性”。他将数字平台放置于全球资本主义经济史中进行比较,认为互联网的市场制度在短短 30 年内涵盖了过去 3000 年的发展史——从熟人的偶发贸易发展到半计划半自由的市场体制。

在互联网诞生之初,远程交易是发生在一些少数交流网站和论坛上的点对点(person to person)交易,相比于数字平台,这些交流网站或论坛没有正式的制度规则和中心化的管理体系,信息相对公开透明,是互联网先驱理

想中的“自由网络空间”。但是,这种自由也意味着人们需要解决霍布斯的经典问题:当人们在“自然状态”,即没有规则、限制、法规等情况下行动时所引发的暴力与欺诈。对于崇尚自由主义的互联网先驱们而言,他们尝试使用道德准则的约束和在线名誉系统两种非正式的解决方案。但是,这两种非正式制度都没有真正解决早期的市场乱象。对于互惠互利的道德准则而言,这种君子协定并不适用于大规模的远程交易。在用户规模不断扩大的网络论坛中,人们不必同一个人产生长期的,甚至于第二次交易,因此“背叛”反而是最大化收益的选择。对于在线名誉系统而言,名誉与自身账户挂钩,因而用户只需重新注册一个新账户以洗白过去的劣迹、舍弃过去的“身份”即可以极低的成本实施欺诈。

交易欺诈等市场乱象的频频出现,不仅使得电子商务市场变得臭名昭著,还使得电子交易市场陷入增长停滞。面对市场失灵的窘境,网站管理者开始介入干预,通过建立一系列互补的正式制度,实现了从网站论坛到中心化管理的数字平台的转变。数字平台为解决早期的市场秩序问题担任制度供给者的角色,它们通过计算机程序和系统建立了一整套互补的市场规则,以帮助网站内部的交易和服务购买活动得以顺利进行。从莱顿维塔的案例中,可以归纳出平台建立的三个正式制度。

第一个是数字身份认证。早期的平台管理者开始要求用户进行“身份注册”,以确保潜在的市场破坏者不会以过低的成本重复注册账户。早期平台借助于国家的力量来确认网民的身份,但随着谷歌、脸书、苹果等数字巨头占据主要市场,这些平台开始将颁布数字身份的权利转移到自己手中,他们无须知道你的真实身份等合法证明,只需有一个“虚拟护照”来给予你畅行平台(间)的权利。

第二个是支付托管系统。与线下“一手交钱一手交货”不同,互联网的远程交易因支付的延时性而存在较大的交易风险,商家没有收到足额的款

项、用户对商品质量不满而要求退款的现象层出不穷。平台管理者设计出了一套资金托管系统,买家需要首先将货款发至平台的托管账户中,卖家才会发出商品,直至买家确认商品无误之后,平台才会将资金支付给商家。随着支付平台的出现,越来越多的日常交易平台将托管系统交予更加安全和稳定的第三方支付平台。

第三个是仲裁制度。电子商务市场规模的扩大,产生了更多需要处理和解决的交易纠纷,早期的平台拥有者也开始雇佣专员来判决、执行市场规则。仲裁制度的建立,推动着平台向企业和组织科层化的过渡。雇佣制的出现,首先使得平台从一到两个人运营的网站转变为一个具有上下层级的组织,为了实现更加可持续的盈利,平台也开始扩展自己的盈利模式,除了收取正常的佣金外,还以资金托管、处理纠纷等市场服务赚取服务费用。正是仲裁制度的建立,使得平台组织开始向复杂结构演化。

数字平台通过满足电子交易问题的制度需求,建立大量互补的正式制度获得了市场权力,开始施行起中心化的统治。早期的数字平台通过相对低廉的服务费和优惠政策吸引用户涌入,但当拥有一定的体量和资产之后,它们不想只是充当被动的基础设施提供者,而是在数据的诱惑下更多地开始为自己谋利。数字平台在市场运营和管理的过程中生成了大量数据,通过对数据的算法统计和分析能够更准确地进行市场定价和精准推送,从而在平台竞争中占据优势;不仅如此,这些平台也在利用算法有意地剥削用户。例如,亚马逊曾为了逼迫发行商以更低的价格和优惠的运输达成交易,通过削弱其推送的算法权重致使这些商家的盈利下降近 40%(莱顿维塔,2024:138),从市场服务的供给者摇身一变为市场的掠夺者。这种封建式的集权统治,标志着数字帝国的初步形成。

（二）权力约束：数字平台的政治制度缺位

莱顿维塔最有创新性的洞见，是分析了云帝国在崛起之后所面临的政治制度与社会制度建设的问题。数字平台的政治制度和社会制度仅仅是一个分析维度的理想型，并不是已完成的客观事实。就莱顿维塔的研究结果而言，市场利益相关者们为了约束数字平台权力而建立的政治制度，建立在市场与技术等长效制约机制失效的基础之上，而自身并未建成一个可持续的权力监督制度；关于帮助和保护电子商务市场中用户与劳动者的社会制度，更是仅发挥着有限作用。但这两种制度在实践上的失灵或缺位并不意味着没有价值。莱顿维塔对数字平台政治与社会制度的梳理，说明了制度并不是一个由数字平台牺牲用户、商家或劳动者等利益而建成的结果，而是一个数字平台权衡各方利益而动态建构的过程，这一历史过程伴随着平台统治者和内部利益主体的博弈。

在说明市场利益相关者所做的政治斗争努力之前，莱顿维塔率先论证了为什么没有产生一个制约数字平台权力的长效机制。第一个可能的权力制约来自市场自身的力量，用户可以自由选择退出平台来逃离平台的剥削和统治。但莱顿维塔提出，数字平台作为一个市场主体（企业）的产品并不是具体的某件商品，而是一套制度规则。制度框架的选择是集体性而非个人性的，一个人不能选择与他人不同的市场交易规则；另一方面，许多细分市场中都至少有一家大型平台企业，但它们之间并非完全是互相替代的关系，大部分用户在这些平台上均有注册，选择退出某一平台只会消耗自己积累的客户积累、声誉等市场资源。第二个可能的权力制约来自技术取代人治。区块链技术作为一个点对点的电子交易系统，将账户核对的责任分散于随机的用户之间，能够取代数字平台在远程交易中的监督作用。但之所以区块链没有创造出一个普遍的、去中心化的市场权力结构，是因为技术虽

然能够实现管理市场的自动化,但无法实现“立法的自动化”。中本聪等区块链发明者并未对区块链本身施加过多的规则,这使得区块链被各种投资者、交易网站、挖矿公司、非法银行等其他利益相关者所掌控,反而催生出一套权力更加错综复杂的“加密统治”。

在市场和技术机制失灵的情况下,莱顿维塔的研究表明,唯有平台上被剥削阶级联合起来的集体行动才能够真正冲击和重塑平台的市场规则。在平台崛起的过程中,不乏下层数字劳工联合起来反抗不公正待遇的案例,但莱顿维塔通过对 Turker 平台上数字工人领袖米兰达案例的分析,认为互联网平台上的数字劳动者在承担着巨大生存压力的情况下,还没有足够的金钱、时间和精力投入到集体运动之中。因此,尽管数字劳工时常提出反抗平台的要求,但缺乏对抗平台的主客观条件。

数字中产阶级则被认为在反抗平台中心化统治中具有变革性力量。莱顿维塔提出,数字中产阶级不仅有相应的资产基础,还拥有较好的教育背景以及广阔的人脉,能够将自己的资源转化为政治力量与平台上层相抗衡。书中介绍了这样一个典型案例:安德鲁·加德基(A. Gazdecki)是自开发软件 Business Apps 的创始人,为了抵制苹果公司颁布的一项限制使用商业化模板创建应用程序的规则,他联合其他应用开发者发起请愿,不仅在媒体上撰写文章,还联合州议员与苹果公司交涉,最终迫使苹果公司做出妥协。正如近现代以来封建领主被其创造的市民阶层的革命摧毁一样,平台所催生的数字中产阶层是引导市场走向更加公平、开放结构的关键。

(三)权力向善:数字平台社会制度的探索

数字平台构建了一个庞大的电子商务市场,深刻改变了社会的就业结构。越来越多的个体依靠平台经济维持生计,出现了数字零工、外卖骑手、滴滴司机、自媒体运营者等新兴职业。在中国,这些新兴职业与体制内和企

业单位的劳动者相比,以个体而非集体单位缴纳社保,由于经济压力、社保认知等原因,整体参保率并不高。这些群体大多生活在没有安全网制度兜底的真空之中,抗风险能力较低。既然数字平台作为电子商务市场的统治者,并依靠这些群体的数字劳动而赚取利润,莱顿维塔认为数字平台在建立保护和帮助人民的社会制度方面做出了一些尝试。莱顿维塔主要梳理了两类数字平台建立的社会安全网。

第一类数字安全网是由数字平台运营的各种慈善众筹平台,如硅谷的Go Fund Me平台和中国的水滴筹平台。在众筹平台出现之前,社会也有非正式的安全网。当人们因生活变故需要救助时,这种非正式的安全网依托于个体的强关系社会网络,从亲缘关系、朋友邻里等人脉资源中获得帮助。在线众筹平台正在成为人们非正式地筹集应急资金度过困境的重要渠道,受益者通过社交媒体分享的方式,利用滚雪球式转发将筹资范围从强关系扩展至弱关系,甚至是无关系网络。第二类是由数字平台开发的医疗保险项目,如亚马逊医疗(Amazon Care)。美国的许多大平台对私人医疗领域加大了投资,但其实践成效相对有限,更多停留在探索阶段。

莱顿维塔认为,社会制度体现了数字平台在公平与效率之间所做出的权衡,也体现出私利性和公共性之间微妙的角色矛盾。作为一个市场公共事务承担者的类国家主体,数字平台会生产出除市场制度之外的其他类型的公共产品,甚至类似于福利国家制度的东西。但就像俾斯麦为了保护本国的人力资本而对医疗保险、意外险和养老险全面立法一样,建立社会制度的动机可能出于工具理性。驱动数字平台建立社会制度的也并非出于人道关怀,而是企业的私利性。它们为了保有优质的现场工作者——物流人员、内容审查员、数据标签员等,而推广建立数字医疗保险项目。尽管出于竞争或营利的目的,但建立社会保障制度的尝试仍然对于数字劳工和消费者的生存保障权有重要意义。

四、对既有研究的推进：数字宪治主义作为平台治理的新方案

莱顿维塔的云帝国理论不仅对平台帝国主义脉络有所发展，对于整个平台资本主义的批判理论而言也同样具有自身特色和价值。这一价值尤其在与近年西方左翼学者提出的新思潮的对比中彰显出来，其中技术封建主义与数字社会主义这两个平台资本主义的批判理论尤为典型。

技术封建主义是指这样一种思潮，即认为当今数字平台统治下的资本主义在许多方面遵循着封建主义逻辑（涂良川、潘依琳，2023）。技术封建主义对平台政治的批判与莱顿维塔的云帝国思想如出一辙，两者都认为数字平台以中心化的权力结构进行资本运作和利益分配，平台“倒退”的政治逻辑与“进步”的经济行为表现出违和的异步性。两者的理论亲和性尤其体现在对“云”与“租”的理解上。技术封建主义将平台管辖的云空间称为“领云”，以对应国家中“领土”的概念（蓝江，2025）。任何想要进入云封地并进行市场活动的厂商，都需要向作为云贵族的平台缴纳“云地租”。正如索伦森（Sørensen，1996）将“租”视为占据支配地位和资源的阶级所获取的非生产性收入（吕鹏，2022），云贵族这样的科技垄断巨头也是通过云租金的资本积累而构建不平等的社会等级体系的。

但是，技术封建主义过于夸大了平台与封建国家的似然性，仅凭对两者生产方式、社会关系和政治逻辑流于表象的分析就否认了平台模式的资本主义本质，这一观点受到了广泛批评（吴鑫，2023；周亦垚、蓝江，2024）。莱顿维塔的理论在避免陷入历史谬误的同时，又推进了技术封建主义对云领地社会阶层及其权力关系的新理解。与技术封建主义将云领地的社会阶层划分为“云贵族—附庸资产阶级—云农奴”不同，莱顿维塔一方面认为在资

产阶级与云农奴之间还存在“卑微的数字劳动者”,即具有反抗意识和行动但缺乏知识和资源的数字劳工;另一方面又重新审视了资产阶级与平台的权力关系。他不仅否认云领地中缴纳租金的资产阶级附庸于平台贵族,还认为平台经济壮大了这一新市民阶层的力量。应用开发商、在线商人、自由职业者、网络名流等数字中产阶级得益于平台经济而崛起(莱顿维塔,2024:277),并成为反抗贵族权力的坚实基础。因此,相较于技术封建主义的片面分析,莱顿维塔在详细分析了历史资料的基础上认为,市场利益相关者对数字平台权力具有动态建构的力量。

此外,相较于滑向右翼话语体系的技术封建主义,同样为西方左翼学者所提出的数字社会主义持有激进的改良主义立场。在探寻平台权力制约的实践方案层面,莱顿维塔与数字社会主义有着共同的理论目标。莱顿维塔认同数字社会主义对平台资本主义改良的基本思路,如将平台权力分散化、归还属于大众的数字权利等,两者的共性尤其表现在对社会主义经济计算的讨论上。莱顿维塔在分析苏联采用计算机技术进行市场调控的历史案例时,认为苏联经济的停滞并不是计划经济体制本身导致的,计划体制本身不是问题,问题在于不具备处理海量数据和进行算法决策的基础设施(莱顿维塔,2024:114)。相反,计划体制是具备强大的算力和数据处理能力的数字平台迅速崛起的关键。这一观点和数字社会主义重拾算法模型替代市场竞争和价格体系的理论资源具有相同立场。

但莱顿维塔基于对数字平台政治制度的研究,认为数字社会主义的具体方案具有内在局限性。其一是依靠区块链技术难以实现真正的分布式自治。区块链社会主义倡导依靠去中心化的自治组织来重新设计基于代码治理的货币系统,或试图建立一个无货币的经济系统(潘娜,2024)。正如前文所介绍的,区块链确实是具有革命性的规则执行系统,但莱顿维塔认为立法和管理是一体两面的,依靠技术自动执行约束权力的规则是不够的,治理

难题的关键在于如何立法(莱顿维塔,2024:176)。其二是数字劳工自治难以动摇平台统治。在数字社会主义思潮中,一些学者从自治主体论的角度认为数字工人是对抗和变革平台资本主义的力量。代表性学者斯蒂安·福克斯提出了数字工人集体运动的一系列方式,如替代性设计数媒技术、数字大罢工等(孙昊鹏、陈子昂、彭敏,2024)。莱顿维塔也讨论了用户合作建立自己平台的案例,但这些平台合作社最终都走向了失败。尽管这些数字劳工的联合抗争切实深入到了"立法权力的分散",但始终没有能够威胁到平台巨头的统治。

与数字社会主义的激进立场相比,莱顿维塔并不支持颠覆平台现有制度,而是持有"数字宪治主义"的立场,提出创立以宪法法治为基础的合作式民主。数字宪法并不是传统意义上由权力机构(如国家或立法机关)颁布的宪法(Redeker, Gill & Gasser, 2015),其内涵更接近一些学者所描述的"小宪法",即反映基本价值、在社会中发挥"宪法功能"的规则(Bradford, 2023)。莱顿维塔如此重视数字宪法的原因,一方面源于他基于制度经济学尊重平台经济运行的制度基础,认为平台经济是数字劳工和数字中产阶级赖以生存的环境,因此在平台原生制度的基础上进行民主改良;另一方面源于他重视在平台立法的源头规范化规则建立的过程,保有平台制度就意味着承认平台在数字经济中的支配性地位,因此与其消解平台的治理权力,不如分散平台的立法权力。

正如阿莫雷蒂(Amoretti, 2009)将数字宪法定义为数字技术激活的一系列历史构成的实践,莱顿维塔的数字宪治思想也是一个各种利益主体协调、博弈和框定中心化权力的实践过程。根据莱顿维塔的阐述,数字宪治的建立包含以下几个关键环节。第一,政策制定者引导平台上的各社会阶层团结起来结成联盟。民族国家作为领土主权的主体,需要鼓励和引领平台内部和外部政治联盟的成立,包括平台公司的内部员工,以及平台外部的消

费者、商会、工会,从而壮大各个阶层的集体性力量。第二,将平台的服务条款上升到“宪法”地位。平台需要在各方的监督下修改和公布正式的服务条款,作为一切权力执行和决策的合法性来源。第三,建立制度化的平台司法监督和审查制度。督促平台建立更加便捷和有效的投诉处理程序,并将自己的司法判决数据化和公开,接受政府和社会的监督。

总之,在与技术封建主义和数字社会主义的比较中,可以发现莱顿维塔在批判平台资本主义和提出替代性解决方案上的理论贡献。数字宪治主义提倡不从根本上触动平台原生制度,鼓励推动平台从人治走向法治。在政府和数字劳工、数字中产阶级等多方利益联盟的监督下颁布平台宪法,将平台政治透明化。

五、结论:云帝国的政治经济学批判

莱顿维塔在制度经济学的视角下提出了云帝国理论。他通过对云帝国崛起的制度历史分析,来说明数字平台是如何建构起现有帝国秩序的。在这一过程中,数字平台通过制度供给解决了互联网电子市场早期的一系列协调问题,确立起自己在电子商务领域的中心统治地位。但莱顿维塔的考察表明,云帝国的统治并不是牢不可破的,数字平台需要在权衡各利益主体的基础上重塑市场制度,其中市场利益相关者(尤其是数字中产阶级)在制度变革中发挥着关键性作用。最后,莱顿维塔对数字平台社会制度的简要分析,说明了数字平台可能会出于私利而生产市场制度之外的公共产品,如数字安全网这一类似于福利国家的社会保障。

《云端帝国》作为一本严谨、科学的学术专著,从制度经济学的视角推动了对平台与国家关系、平台权力以及治理的新理解。首先,与平台帝国主义的脉络相比,莱顿维塔对平台政治的批判建立在客观而详尽的历史研究

中,制度经济学和经济社会学的知识背景,使得他能够冷静客观地分析数字平台帝国化的经济制度动因。其次,虽然表面上他将平台类比为国家,但他并不像技术封建主义一样完全将平台政治理解为封建逻辑;相反,莱顿维塔将数字平台放置于全球资本主义经济史中,审视平台与民族国家的权力互动过程,强调类国家的权力中心化运作反而是推动平台资本化的原因。最后,莱顿维塔的理论否定了数字社会主义倡导的技术的分布式治理和劳工自治,强调问题的关键在于分散化平台的立法和司法权,采用数字宪治的方案以解决现状。因此,可以认为莱顿维塔的云帝国理论是尊重经济和制度运行规律、具有独特价值的平台政治经济学批判。

除了学术贡献,《云端帝国》也是一本面向更广泛受众的出色读物。这一方面得益于莱顿维塔尽可能少地采用抽象而复杂的理论(甚至在全书中没有提到制度经济学的理论传统),仅以“制度”为核心构建框架,让读者轻松地理解数字平台帝国崛起的历史动因;另一方面,莱顿维塔善用纵向比喻来说明数字平台的技术治理与国家统治之间的相似性,这种巧妙的技法增加了可读性。例如,在第五章花费大量篇幅介绍苏联的计划经济依靠“数据库”和“计算”来运转,以此说明为什么数字平台在经济制度上也是一种“中央计划”;在第七章详细介绍了雅典采用“分配机器”来随机抽签选举城邦管理者的做法,以此说明区块链技术的内在机制;在第十章介绍俾斯麦推动建立社会安全网的故事,来说明数字平台建立社会制度是由工具理性和追逐利益而驱动的。总之,这些众所周知的历史典故帮助人们更好地理解了为什么数字平台的内部治理遵循了传统的治国之道。

同时,《云端帝国》也存在一些不足。汇编类的章节编排能够惠及更广泛的读者,但也因此缺乏一个统领性的理论框架。莱顿维塔对制度概念的定义清晰而明确,但将其划分为经济制度、政治制度和社会制度的做法也引起了一些诸如概念主体不清、制度间关系不明晰等问题。此外,莱顿维塔着

重于探讨经济与政治制度,而对于社会制度的探讨稍显薄弱。仅用一个章节以慈善众筹平台(以及少量的技能培训制度)为例,难以支撑起数字平台为保护和保障人民建立起“制度”的概念假设。

参考文献

蔡跃洲、王麒植、钟洲,2024,《线上排他行为、阶段性特征与数字平台治理:三方动态博弈分析》,《经济研究》第 5 期。

杜巧玲,2024,《数字平台与平台资本积累的政治经济学批判》,《学术交流》第 2 期。

莱顿维塔,2024,《云端帝国:数字平台如何重塑经济与世界》,吴晓昆译,北京:中译出版社。

蓝江,2025,《什么是技术封建主义?——当代西方技术封建主义的政治经济学批判》,《西北师大学报(社会科学版)》第 1 期。

罗必良,2005,《新制度经济学》,山西:山西经济出版社。

吕鹏,2022,《全体人民共同富裕与社会学的使命担当》,《社会学评论》第 6 期。

吕鹏、周旅军、范晓光,2022,《平台治理场域与社会学参与》,《社会学研究》第 3 期。

马俊峰、赵海蕴,2024,《平台资本主义对“玩劳动”的三重剥削探赜》,《贵州大学学报(社会科学版)》第 3 期。

诺斯,1995,《制度变迁理论纲要》,《改革》第 3 期。

潘娜,2024,《数字社会主义:问题域与理论流脉》,《开放时代》第 5 期。

邵占鹏,2024,《数字平台的三重属性与角色冲突》,《浙江学刊》第 5 期。

斯尔尼塞克,2018,《平台资本主义》,程水英译,广州:广东人民出版社。

孙昊鹏、陈子昂、彭敏,2024,《国外数字社会主义探索及启示》,《数字经济》第 6 期。

涂良川,2022,《平台资本主义技术逻辑的政治叙事》,《南京社会科学》第 2 期。

涂良川、潘依林,2023,《技术封建主义:科技进步的政治退步——当代西方左翼“技术封建理性批判”刍议》,《世界社会科学》第 2 期。

王沛楠、彭修彬,2023,《西方数字平台的内容治理:理解互联网国际传播中的平台权力》,《现代传播(中国传媒大学学报)》第 4 期。

王志鹏、张祥建、涂景一,2016,《大数据时代平台权力的扩张与异化》,《江西社会科学》第

5 期。

温旭,2024,《从数字帝国主义到数字帝国:数字资本主义的全球化逻辑》,《理论与改革》第 1 期。

吴鑫,2023,《马克思主义视域下“技术封建主义”思潮批判》,《社会主义研究》第 3 期。

谢尧雯,2021,《网络平台差别化定价的规制路径选择——以数字信任维系为核心》,《行政法学研究》第 5 期。

严运楼,2020,《当代西方左翼关于数字帝国主义批判研究》,《毛泽东邓小平理论研究》第 6 期。

张谦、李冰晶,2021,《数字平台捆绑销售策略的垄断动机研究:基于交叉网络外部性视角》,《管理学刊》第 2 期。

张雅俊,2023,《数字平台算法歧视定价:私利性与公共性冲突及其监管因应》,《价格理论与实践》第 3 期。

郑吉伟、初传凯,2024,《数字平台参与剩余价值分配的政治经济学分析》,《经济发展研究》第 1 期。

周亦垚、蓝江,2024,《资本主义的雅努斯:资本抑或云地租——技术封建主义思潮对〈资本论〉的误读》,《探索与争鸣》第 8 期。

Amoretti, F. 2009, *Electronic Constitution: Social, Cultural, and Political Implications*, [S. l.]: IGI Global.

Bradford, A., 2023, “Europe’s Digital Constitution.” *Virginia Journal of International Law* 64 (1).

Couldry, N. & U. Mejias 2019, *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*, Stanford: Stanford University Press.

Dommett, K. 2021, “The Inter-Institutional Impact of Digital Platform Companies on Democracy: A Case Study of the UK Media’s Digital Campaigning Coverage.” *New Media & Society* 25(10).

Ergen, T., L. Antonucci & M. Arslan et al. 2024. “On Vili Lehdonvirta’s *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control*, Cam-

bridge, MA., MIT Press, 2022." *Socio-Economic Review* 22(4).

Gawer, A. 2022, "Digital Platforms and Ecosystems: Remarks on the Dominant Organizational Forms of the Digital Age." *Innovation* 24(1).

Jin, D. 2015, *Digital Platforms, Imperialism and Political Culture*, New York: Routledge.

Lin, Z. 2021, "Commercialization of Creative Videos in China in the Digital Platform Age." *Television & New Media* 22 (6).

Muldoon, J. 2022. "Book Review: *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control* by Vili Lehdonvirta." https://blogs.lse.ac.uk/usappblog/2022/09/11/book-review-cloud-empires-how-digital-platforms-are-overtaking-the-state-and-how-we-can-regain-control-by-vili-lehdonvirta.

Rani, U. & M. Furrer 2021, "Digital Labour Platforms and New Forms of Flexible Work in Developing Countries: Algorithmic Management of Work and Workers." *Competition & Changee* 25(2).

Redeker, D., L. Gill & U. Gasser 2015. "Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights." *International Communication Gazette* 80(4).

Sørensen, A. 1996, "The Structural Basis of Social Inequality." *American Journal of Sociology* 101(5).

技术对社会的再结构

——从卡斯特的网络社会到智能社会

叶莹菲*

摘要:20世纪末,卡斯特在《网络社会的崛起》中论述了技术对社会形态的再结构。本文基于卡斯特网络社会理论的研究视域,以从网络社会到人工智能社会的发展为经,聚焦生产力与生产关系和就业结构的转型,试图比对、勾勒出当前智能社会的图景。在智能社会中,以人工智能为核心的生产力具有高度集中的天然特性,掌握算力、算法和数据的主体将获得前所未有之垄断的可能性。此外,本文深化了对人工智能是否将取代人类劳动这一现代性问题的认识。目前,人工智能技术的扩散与就业水平的演变还没有显示出系统性的结构关系,就业被取代和被创造并存。

关键词:卡斯特　网络社会　人工智能　就业结构

一、引言

20世纪90年代初期,互联网开始在全球范围内蔓延。自此以后,网络的应用建构了社会新形态,渗透到人类生产生活的各个领域。随着数字技术和移动互联的普及,人类文明进入了数字社会时代。2010年以后,以数据、算法、算力为核心的人工智能技术呈现爆发式发展,全球进入迈向智能

* 叶莹菲,中国社会科学院大学社会与民族学院。

社会的新阶段。在近 60 年信息、数字和人工智能技术的发展过程中,技术参与了社会演变的全过程,型构了社会新形态的物质基础,与组织、制度、国家、文化一道塑造了新的社会结构。

20 世纪末,卡斯特(M. Castells)捕捉到了社会因信息技术革命而发生的结构大转型。《信息时代三部曲:经济、社会与文化》成为网络社会学和数字社会学的先驱之作,并被吉登斯(A. Giddens)誉为能与马克斯·韦伯《经济与社会》相媲美的著作。其中,《网络社会的崛起》一书分别从新经济的形态、生产力与生产关系、企业组织的网络化、就业结构的再造、时间和空间、沟通象征等方面,论述了信息主义对资本主义再结构的过程。

在卡斯特看来,网络社会起源于 20 世纪 70 年代左右的信息技术革命,根植于信息技术范式和网络化的逻辑。信息主义将资本主义再结构,改变了生产、经验、权力和文化过程中的操作与结果,渗透到生产和生活的各个领域,形成新的社会形态,即网络社会,而网络社会中的新经济具有全球化、信息化和网络化等特点。企业组织以不同几何形态的网络连接起来;在技术、制度和组织等因素的互动下,就业结构将发生转变,弹性工作者愈发涌现;时空被信息技术重塑,形成流动的空间和无时间之时间。

目前,以人工智能、大数据、云计算为核心技术的第四次工业革命正在展开,人类社会开始迈向智能时代,智能社会成为学界关注的热点。以人工智能技术为代表的技术变革,将会带动生产方式、生活方式、社会治理、人际关系等方面的变化。本文中的“智能社会”,指的是以人工智能技术为核心的社会形态,如同数字技术之于数字社会,信息技术之于网络社会。

本文受《网络社会的崛起》的启发,从卡斯特庞大的叙事框架中摘取了部分研究视域,聚焦生产力与生产关系、就业结构转型这两部分,试图在《网络社会的崛起》出版 28 年后,回应、比对、勾勒出当前智能社会的图景。之所以框定这两个方面,是因为从社会结构转型角度来看,生产力和生产关

系依旧是基础,起着决定性作用(杨善华,2016:103),并且影响着全球经济的生产、分配、交换、消费以及权力格局。此外,劳动过程与就业结构是社会结构的重要组成部分,直接关系到社会的经济运作和生计。本文试图通过上述分析,对人工智能是否将取代人类劳动的社会关切做出回应。

二、技术基础及与社会的关系

同哈维·布鲁克斯(H. Brooks)、丹尼尔·贝尔(D. Bell)(1984)的看法一致,卡斯特认为技术是指运用科学知识,以一种可以复制的方式,来详述做事情的方法(卡斯特,2001:34)。在他的概念中,信息技术除了传统的微电子、电脑(硬件和软件)、电信、广播以及光电等汇合而成的整套技术,还包括与前者汇聚互动的遗传工程。

通过梳理技术发展的脉络,我们可以大概了解社会发展所立足的技术物质基础,及其带给社会发展的可能性。从技术革命的历史序列来看,信息技术革命的核心是处理程式电子计算机和电晶体,后者诞生于 20 世纪 70 年代(卡斯特,2001:46)。三个主要技术领域,即微电子学、电脑与电信的创新与突破,及其相互之间的互动,构成了技术史的基础。具体而言,电晶体、微电子、晶片、电脑、微处理器、电子交换器、光纤与镭射传输、数码封包传输、无线电波段的运用、同轴电缆技术等,都为信息革命提供了技术物质基础。

20 世纪 80 年代,为防止苏联破坏美国通信网,由美国国防部先进研究计划局资助的互联网进入大众视野。1973 年封包通信技术促进了区域网络的出现,使得电脑之间可以彼此对话。1978 年,“骇客”让电脑可以通过电话线彼此连接。个人电脑与网络沟通能力的出现,刺激了电子布告栏系统(BBS)的发展,虚拟社区由此诞生(Rheingold, 1993)。1990 年,位于日

内瓦的欧洲核子研究中心发明了万维网,浏览者通过搜寻系统可以找到他们想要的信息。1994 年,随着新的搜索引擎的快速发展,整个世界开始拥抱互联网。正是这个时期,互联网介入商业服务,为数字社会的发展集齐了基本条件——终端(计算机)、网络和浏览器(网络人机界面)——行动者从此有了实现互联互通的机会,互联网的社会化应用真正出现了(邱泽奇,2022)。此后,移动电话的普及和革新,使得生产生活的各个领域,如工作、通信、娱乐、教育、交通、购物等,被技术深度嵌入。

人工智能的技术历史基于信息革命的物质基础,却发展、分化出了另一条科技路线,并成为驱动第四次工业革命的核心力量(国务院发展研究中心国际技术经济研究所,2019:6)。“人工智能”这一概念于 1955 年达特茅斯会议上被约翰·麦卡锡(J. McCarthy)首次提出,这次会议的主题是让机器使用语言,形成抽象与概念,解决目前只有人类才能解决的问题,以及让机器实现不断的自我提升。

人工智能技术的发展大致经历了三个阶段:第一阶段是 20 世纪六七十年代,人工智能力图通过计算机来实现机器化的逻辑推理证明,但最终难以实现;第二阶段是 20 世纪 70—90 年代,这时计算机的能力已有了长足进步,因此试图通过建立基于计算机的专家系统来解决问题,但是由于数据较少且太局限于经验知识和规则,难以构筑有效的系统;第三阶段从 20 世纪 90 年代末开始,神经网络等方向的研究被引入,人工智能技术才逐渐步入快速发展期。2010 年后,人工智能技术的发展进入快车道,机器深度学习复兴,神经网络在图像识别、自然语言处理等领域取得了突破。随着大数据时代的到来,人工智能技术开始向许多领域渗透,包括医疗、自动驾驶、语音识别、机器翻译等。算法、算力(芯片)和数据成为人工智能三要素,大数据处理推动了算力的发展,算力提升了算法的探索空间,三者相辅相成。2017 年,由谷歌旗下 DeepMind 公司开发的人工智能程序 A1phazero 击败了当时

世界上最强大的国际象棋程序 Stockfish;2023 年,生成式人工智能取得了惊人的进步,OpenAI 公司旗下的 ChatGPT4. 0 和 Sora 先后发布,意味着通用人工智能(AGI)具有巨大潜力和无限前景。

在某种程度上,20 世纪 70 年代构成一个系统的新技术,乃是 80 年代社会经济再结构的基础;而这些技术在 80 年代的应用,则大体上决定了其在 90 年代的应用和轨迹;以此类推。在不同的技术阶段,因技术原理不尽相同,产品和应用形态有很大差异,其影响社会结构的方式也存在复杂的差异;同时,社会中的主体也在适应、调整、更新原有的社会结构。① 冒着简单化的风险,沿着技术和历史发展的轨迹,社会的演变大致经历了从网络社会到数字社会再到智能社会的过程。尽管三者具有一定的演变顺序,但相互重叠且各有侧重,其所仰赖的技术基础相互渗透并勾连。其中,网络社会以信息技术、通信技术为核心,主要指互联网的应用对社会造成的变革,卡斯特和其他学者也用“信息社会”一词指代网络社会;数字社会是由数字化技术推动的,在大数据、人工智能等技术基础上形成的社会系统,可以从数字经济、数字商业、数字生活等层面来研究数字社会(陈刚、谢佩宏,2020);智能社会是基于人工智能技术(包括算法、算力和大数据)的社会新形态。

但需要注意的是,技术并不决定社会(Marx & Smith, 1991)。20 世纪 40—60 年代的军事资金和市场,20 世纪 60 年代起源于美国硅谷的自由文化、创新和企业精神,都是技术形成的重要参与要素。一旦新技术扩散开来,便为不同国家、文化、组织和目标所挪用,生成各种应用,并反馈迭代技术创新。这不仅表现为经济与社会的嵌入(波兰尼,2020),更可延展为政治、经济、社会和技术的相互嵌入。要了解从网络社会到智能社会的变迁,有必要重视两股力量:一是新信息技术的发展,二是旧社会尝试利用技术力

① 这印证了吉登斯(2016)的结构二重性理论,即社会结构同时具有客观约制性和主观创造性。

量的权力主体。在卡斯特看来,网络社会中的技术的真正主体和推动者是国家和机构(卡斯特,2001:54),经济增长理论化的工作应该围绕技术变迁、公司能力与国家机构三者之间的关系而展开(Nelson, 1994)。对企业而言,获利力的增长推动其发展;对政治机构而言,它受到一组更广泛价值与利益的塑造;而经济领域倾向最大化其经济组成部分的竞争力。国家既可以促进技术发展,也可以窒息其发展。

三、生产力与生产关系的重构:从网络社会到智能社会

(一) 网络社会的生产力与生产关系

在《资本论》中,马克思对生产力和生产关系这两个概念内涵的论述要复杂得多。他将生产力区分为具体生产力(马克思、恩格斯,2016a:59)和一般生产力、劳动的自然生产力(马克思、恩格斯,2016a:56—57)和劳动的社会生产力(马克思、恩格斯,2016b:994)、资本的生产力和劳动生产力(马克思、恩格斯,2016a:53)、物质生产力和精神生产力(马克思、恩格斯,2016b:279)等。在生产关系方面,马克思讨论了以物为中介的生产关系,这些物包括商品、货币、资本、价值、利润、地租、剩余价值等(马克思,2018)。

本文以这两个基本概念为分析工具,并以学者以往的简要定义为研究切入点。本文所谓的生产力是指"生产物质资料的能力或力量。也就是说,撇开生产中人与人的社会关系的一面,一切物质资料生产中所必要的因素,都属于生产力的内容"(卫兴华、李凯明,1980),其表征包括构成生产力系统的要素,如劳动者、劳动资料、劳动对象(李达,1958),以及构成生产力系统的各类要素或成分之间的结构(赵家祥,1989)。马克思所使用的生产关系概念,主要指人们在直接生产过程中的关系,并不包括交换关系、分配

关系、消费关系等。赵家祥等人(赵家祥、聂锦芳、张立波,2003:307—308)称之为狭义生产关系,并将这种生产关系扩展至交换关系、分配关系、消费关系,称之为广义生产关系,还将生产关系定义为人们在物质资料生产过程中结成的经济关系,包括生产资料所有制形式、各种社会集团在生产中的地位以及其相互关系、产品分配形式三项内容。

延续马克思主义对生产力和生产关系叙述的路径,在网络社会中,卡斯特认为生产力的根源是获取技术知识的能力,犹如劳动和自然资源(土地)之于农业发展方式,新能源及其使用作用于工业社会。网络社会具有信息化的特征,而信息化是指"该经济体内单位或作用者(不论是公司、区域或国家)的生产力与竞争力,基本上是看它们能否有效生产、处理及应用以知识为基础的信息而定"(卡斯特,2001:91)。

那么,以知识为基础的生产力是信息化经济所特有的吗(卡斯特,2001:95)?特别是与工业经济相比,它的特殊性何在?要回答这个问题,卡斯特(2001:98)首先比较了先进市场经济体系(G7 和 OECD)1960—1993 年生产力的增长趋势。从统计数据来看,他发现经济增长之"工业化"与"信息化"体制之间,并无多大的差异。他认为,这首先是因为技术和生产力的发展存在时间差;其次,由于统计上的测量不合时宜,无法将技术(知识)作用于生产力的结果表现在服务业上。1999 年后,美国商业部改革了统计分类,将企业软件上的花费计入投资,表现在 GDP 中。于是,1995—1999 年的生产力增长率攀升到了 2.6%(Uchitelle, 1999);1993—2007 年,仅工业机器人自动化就使 17 个国家的劳动生产率增长了 0.36 个百分点(Lee, 2016)。从美国 1959—1999 年的数据(卡斯特,2001:108)来看,生产力增长率的提高主要源自电脑能力、通信与信息技术的大幅度改善(Greenspan, 1998)。

然而,以知识为基础的生产力同样存在于农业和工业领域吗?例如,家庭成员需要掌握农业知识才能够获得好收成,但这些知识和技能无法测量,

更无法测算总量;同样,在工业时代,以知识为基础的信息最终转化为机械和电气设备等生产工具,也同样促进了生产力的质的提高。1999 年,美国商务部将企业软件花费计入 GDP,最终使得知识信息的一部分得到了量化表达;类比工业时代和信息时代,以知识为基础的生产力的差别,则需要使用其他工具进行测量,如将“机器投资”类比于“企业软件投资”。因此,以知识为基础的生产力是信息化经济的特征值得商榷,改为以信息技术知识为基础的生产力或许更为妥帖,因为其根本的差别是知识的类别不同:农业时代需要种植、畜牧、病虫防害、灌溉等知识,工业时代需要能源、化工、机械、电力等知识,信息时代则需要以信息技术为基础的知识。

就生产关系而言,卡斯特认为生产关系和社会阶级都由生产方式界定,生产方式由“占用、分配与使用剩余”的规则决定,而社会结构与生产过程决定了这些规则。回顾网络社会崛起的历史,生产力的掌权版图如何?“占用、分配与使用剩余”的规则又如何?卡斯特认为,20 世纪有两种生产方式,一是资本主义,二是国家主义。前者,生产者与生产工具相分离,资本家控制、占有、分配生产工具和剩余价值,以利润最大化为目标;后者由国家掌权者控制剩余价值,以权力最大化为目标。同样,生产力决定了生产的社会关系和生产方式,同时决定了谁占有剩余价值。

从企业组织形式来看生产关系,信息时代的企业组织在生产过程中表现出了全球化特性。以资本主义国家为例,企业为了增加利润,除降低生产成本、发展生产力、加速资本周转这三种方式外,拓展市场也能带来更多收益。因此,20 世纪后半叶,全球化趋势愈加普遍——不仅包括市场的全球化,还包括生产过程的全球化。卡斯特(2001:91)将全球化描述为网络社会新经济的一个主要特点,全球化是指生产、消费与流通等核心活动及其组成元素(资本、劳动、原料、管理、信息、技术、市场)在全球尺度上组织起来;若非直接组织起来,则通过经济作用者之间的网络来组织起来。

在全球化进程中,有两个产业是直接受益者:一是信息科技业,二是金融业。20 世纪末,首先,在国家的全球化战略下,贸易和外国直接投资(FDI)大幅扩张(卡斯特,2001:112),成为全球经济增长的引擎;其次,各国市场逐渐解除了管制,使全球市场流动更加自由;最后,新信息科技不仅为全球互联互通提供了技术基础,还提高了资本的移动能力。于是,信息科技业(如微电子业、微电脑业、电信业)与金融业成为这种再结构的直接受惠者(Schiller, 1999),通过伸展全球性的触角、整合市场,其获利能力大幅提升。操纵这些主导产业的,是以美国为首的发达国家。对于政府而言,其竞争力表现为赢得市场占有率的能力,并以此作为政府正当化的主要力量(Zysman & Tyson, 1983)。如此,全球性资本主义公司及其附属网络形成网络,其主导产业突飞猛进地发展,而网络之外的企业则开始衰退。这就引出了卡斯特意义上新经济体的第三个特征——网络化。

网络化是网络社会中企业组织在生产过程中表现出的另一重要特征。网络化指生产力的提升与竞争的持续均在全球企业网络之间的互动中实现(卡斯特,2001:91),这首先体现为企业间网络的形成,如上下游市场与合作关系的网络化。卡斯特列举了网络化的几种典型模式:小企业间的网络,尽管中小企业常受大公司支配,但它们也会与大公司或其他中小企业建立网络关系,以增强自身竞争力;“贝内顿模型”(Benetton Model)(Harrison, 1994)以财务控制的垂直结构为核心,同时在全球范围内构建水平式的企业网络;公司战略联盟,这种模式在高科技产业尤为常见,合作内容涵盖研发、营销渠道、资金与技术交易等方面,大型企业不再是自给自足的系统,而与众多外包商和协作厂紧密相连;全球网络化企业模型,如思科公司将客户、供应商、合作伙伴和员工整合到网络化逻辑中,并通过自动化实现大部分互动工作,其核心运营主要依托公司网站。除了企业间的网络化,网络化还体现在企业组织内部的结构上。与传统的垂直科层制相比,现代企业更倾向

于采用水平式网络结构。这种结构要求企业本身成为一个网络,激活内部每个元素,赋予各单元以自主权,并在整体战略下进行竞争(Goodman & Sproull, 1990; Imai, 1990)。

在卡斯特的理论框架中,分散的计算机通过互联实现了重新组织的可能。网络企业[①]的兴起,标志着经济组织基本单位的根本转变。历史上头一次,经济组织的基本单位既非单一主体(如企业家或家庭企业),亦非集体主体(如资产阶级、公司或国家),而是由多个主体和组织构成的网络,这一网络通过不断调整以适应环境与市场结构的变化(卡斯特,2001:245)。换言之,企业及其产业关系(包括产业上下游)以及合作网络,共同构成了网络社会中竞争的主体。

综上所述,在网络社会中,先进生产力主要集中于以知识与信息为核心的高新技术企业,而这些企业又高度集中于发达国家。发达国家(以美国为首)的政府与多国/跨国企业通过全球市场拓展、金融产业布局以及高科技产业的全球网络化,实现了生产、交换、分配与消费过程的全球化。这种全球化不仅巩固了其在全球经济中的主导地位,还确保了其持续的竞争力和影响力。同时,这些发达国家与跨国企业主导了"占用、分配和使用剩余"规则的制定与实施。

因此,国家及其企业在全球范围内的竞争不可避免地推动了全球经济的区段化与网络化。这一过程导致资源、活力与财富集中于特定区域,进而加剧了全球范围内的不平等与社会排斥趋势(卡斯特,2001:156)。例如,高附加值和高科技产品几乎完全由发达经济体主导,多国公司通过跨国生产构建全球生产网络,而网络之外的国家与企业则被边缘化。而且,以发达国家企业利润为导向的选择性科技发展,进一步巩固了这种不平等格局。

---

① 一种特殊的企业形态,其运作机制由多个自主目标系统的部分相互交织构成。网络中的组成元素兼具自主性与依赖性,既独立运作又相互依存(卡斯特,2001:91)。

（二）智能社会：生产力与生产关系的重构

在智能社会中，从劳动主体的角度来看，人工智能技术将延展人类的劳动能力，突破人类劳动在时间和空间上的限制；从劳动资料的角度来看，人工智能领域的劳动资料主要表现为算力与算法，而在更广泛的领域中，人工智能技术则推动了劳动资料的智能化升级；从劳动对象的角度来看，人工智能领域的劳动对象是数据，而在其他领域中，人工智能技术将促进劳动对象的智能化和数字化（王水兴、刘勇，2024）。

与信息技术相比，人工智能技术作为一种新型生产力，对社会发展的影响与以往有何不同？弗雷等学者（Frey & Osborne，2017）指出，人工智能技术对以往的所有模式提出了挑战，因为它使整个就业领域和生产活动同时受到冲击，这种技术将机器的速度与效率同人类的能动性相结合。人工智能是首批真正的“通用机器”，能够在所有行业和各类工作中得到应用，并能随着可用数据的增加而持续学习与改进。

目前，学术界关于人工智能影响生产力的观点存在分歧。一些学者认为，人工智能能够通过多种途径提升生产率，如通过提高预测精度来减少不确定性（Agrawal，Gans & Goldfarb，2019）、实现技术自动重组（Agrawal，McHale & Oettl，2019；Cockburn，Henderson & Stern，2019），而更普遍的则是创新驱动（Bartelsman，Falk & Hagsten et al.，2019）。基于对 5257 家人工智能专利企业的数据分析，达米奥利等学者（Damioli，van Roy & Vertesy，2021）在控制企业非人工智能相关领域专利活动的情况下发现，2000 年至 2016 年间，人工智能专利申请对劳动生产率具有积极且显著的影响，这种影响主要集中于中小企业和服务业。这是因为规模较小、灵活性较高的人工智能专利公司能够更快地在其生产过程中大规模引入人工智能应用程序，从而显著提升生产率。相比之下，大公司可能需要更长时间才能在其价

值链中充分利用该技术。此外,尽管一些企业已经掌握了前沿核心技术,但在追求整体利润最大化的驱动下,它们可能采取缓慢释放技术应用的策略,并通过不断升级技术和应用,让市场分期为其买单。这也是生产率在数字上无法快速提升的另一个原因。

自 20 世纪 70 年代以来,发达经济体的生产率增长持续低迷,即使在大量投资于数字创新的行业中也不例外,这一现象被称为“生产率悖论”(Brynjolfsson, 2010)。这一悖论可能源于技术推动社会经济发展的滞后性,企业需要通过组织重组来提升劳动技能,之后技术效应才能在整个经济中扩散。戈登(Gordon, 2016; Gordon, 2018)指出,生产率放缓可能是永久性的,当前对数字技术或人工智能革命推动生产力增长的预期过于乐观。与美国 20 世纪 20—70 年代电力和内燃机等带来惊人生产率增长的创新相比,人工智能的推动作用相对有限。基于对科技创新人员研发数据的分析,琼斯(Jones, 2009)发现,首次研发成果的推出时间不断延后,研发过程的专业化程度越来越高,且愈加依赖团队协作,因为研究人员需要投入更多时间学习以掌握推动前沿技术的能力。格雷斯等人(Gries & Naudé, 2018)则从另一个角度强调了总需求的潜在作用,特别是自动化和人工智能导致收入与劳动分配下降,加剧了不平等,使得创新剩余难以被广泛获取,这可能会抑制潜能的释放,从而阻碍生产率的提升。

在人工智能社会中,首先,以算法、算力和数据为核心的生产资料仍然掌握在资本家手中。其次,人工智能改变了劳动的组织方式。在成熟的人工智能社会中,低技能的人类劳动力将被大幅替代,劳动组织将倾向于以人工智能为基础的机器人自动化生产模式,生产关系也将随之发生深刻变革。如果说劳资关系是工业时代问题的核心,那么到了智能时代,机器和资本与劳动者之间的关系则可能成为这一时代的核心议题。这是否意味着我们离“无工作社会”更近了一步?

此外,人工智能与资本的结合,将使掌握算力、算法和数据的主体获得前所未有的垄断可能性。纳雷恰尼亚(Narechania, 2021)指出,正如铁路、电话等行业一样,机器学习具有天然的垄断性。人工智能垄断性的形成依赖两个关键要素:一是机器学习的研发、维护和优化需要巨额资金投入;二是机器学习的训练和优化成本具有累加性,其他竞争者无法绕过这些累加性工作。此外,大型公司掌握着机器深度学习算法的核心——数据,通过并购小型创新公司获取新技术和人才,进一步巩固了其垄断地位。同时,这些公司通过制定人工智能领域的技术标准和生态系统,在知识、技术、资本和网络中占据了绝对优势,形成了供给端的高壁垒,使其他竞争者难以替代。在需求层面,人工智能产品与互联网类似,具有显著的"赢者通吃"效应,这是因为用户获取优质产品和较差产品的成本相同甚至更低。

四、劳动与就业的结构性的变迁

(一) 网络社会就业结构的变迁

社会结构变迁最直接的表现之一是就业与就业结构的转型。传统的后工业主义与信息主义认为,新社会的就业结构表现为从商品生产向服务业的转变、管理与专业性职业的崛起、农业与制造业工作的减少,以及多数发达经济体中工作内容的信息化程度逐渐提高(卡斯特,2001:248)。卡斯特则采取了不同的研究取向。基于经验研究,他考察了七大工业国(G7)1920—1990 年就业结构的演变,揭示了以下基本趋势:农业就业逐渐萎缩;传统制造业就业持续衰退;生产者服务和社会服务兴起,前者以企业服务为主,后者以医疗保健服务为核心;服务业活动作为就业来源日趋多样化;管理、专业和技术性职业快速增长;由服务员和销售员构成的"白领"无产阶

级逐渐形成;零售业就业比例显著且相对稳定;职业结构的顶端和底层同时扩展;职业结构随时间推移逐步升级,需要更高技术和高等教育的职业比例增速超过低层次工作的增长。

在普遍的共性之外,这些信息化社会国家还呈现出文化与制度的多样性。卡斯特提出了两种不同的信息化模型假说:第一种是服务业经济模型,以美国、英国和加拿大为代表,其特点是制造业就业比例下降,农业就业几乎消失,服务业内部出现分化,且更重视资本管理服务而非生产者服务;第二种是工业生产模型,以日本和德国为代表,尽管制造业就业比例有所下降,但仍维持在 1/4 左右,且生产者服务的地位远高于金融服务。当不同经济体迅速走向整合与交互渗透时,其就业结构将大致反映出每个国家和地区在全球生产、分配与管理结构中的位置。

在网络社会初期,“无工作社会”的恐惧便已开始蔓延。信息技术在工厂、办公室和服务业中的广泛应用,重新点燃了几个世纪以来工人对机器取代人力的担忧(Allen, 2009)。

然而,从美国和日本的历史经验中可以发现,尽管农业劳动力被替代,传统制造业也经历了与农业相似的命运,但新的工作岗位不断涌现,尤其是在高科技制造业和更为显著的服务业领域(OECD, 1994)。1970—1992 年,美国经济实际增长了 70%,就业增加了 49%;日本经济实际增长了 173%,就业增加了 25%;而欧洲共同体经济增长了 81%,就业仅增加了 9%(Commission of the European Communities, 1994)。卡斯特指出,某些欧洲国家的高失业率主要并非由新技术引入所致,而是由于错误的总体经济政策,如未能有效鼓励私人部门创造就业机会。因此,技术创新与扩散并不会直接创造或破坏工作岗位。基于经济与合作发展组织 21 个国家 1987—1994 年的数据,马丁·卡诺伊认为技术扩散与就业演变之间并无统计上的显著关联(Martin, 2000a)。这项研究与其他分析(Richard, 1994)均表明,

制度差异对失业率有一定影响,但技术水平的影响并未呈现一致模式。1985 年,日本劳动研究局研究了电子技术对汽车、报业等不同产业的影响,发现所有案例中,引入新技术的初衷并非为了减少实际劳动力规模,后续也未导致这一结果(Japan Institute of Labour, 1985)。德国研究与技术部指出,非技术工作在短期内可能被替代,但从长期来看,生产力提升将创造更多就业机会(Schettkat & Wagner, 1990),大多数被取代的工人将重新进入劳动力市场(Miller, 1989)。此外,技术创新对不同产业的影响各异。列维等人(Levy, Bowes & Jondrow, 1984)的研究表明,在铁矿、煤矿和铝矿行业,技术变革增加了产出并提高了就业水平,但在钢铁和汽车行业并未表现出同样的效果。此外,拉斐尔·卡普林斯基(Kaplinsky, 1986; Bessant, 1989; van den Besselaar, 1996)等人提出,必须区分经济的八个不同层次:生产层次、工厂层次、公司层次、产业层次、区域层次、部门层次、国家层次以及世界层次。他们的结论是,宏观与微观的量化研究似乎得出了基本不同的结论。制程与工厂层次的调查通常表明劳动力被显著替代,而国家层次并未出现重大就业问题。

然而,罗马俱乐部提出了相反的观点:技术的发展将对就业产生全新的冲击。这些论点(King, 1991; Aronowitz & Di Fazio, 1994)认为,如果制造业和农业就业沿着相同的轨迹衰退,那么服务业将无法充分吸纳这些劳动力,而且服务业本身也正在经历快速的自动化进程。卡斯特指出,这些观点缺乏严谨的证据支持,而仅仅基于零星的简报和随意的公司案例。相反,西欧、美国和东亚的经济增长为反驳这些观点提供了有力的证据。

总之,信息技术扩散与整体经济就业水平演变之间似乎并不存在系统性的结构关系。尽管某些工作岗位被取代,同时也有新的工作岗位被创造,但得失之间的数量关系因企业、产业部门、区域和国家的不同而有所差异,并取决于竞争力、企业策略、政府政策、制度环境以及在全球经济中的相对

位置。信息技术与就业之间互动的具体结果,取决于宏观经济因素、经济策略以及社会政治背景(Martin, 2000b)。

(二) 智能社会就业结构

随着技术的发展,人类身体的各个部位逐渐被工具替代。从工业时代手和脚的功能被延展与强化,到如今人工智能时代人脑的部分功能被替代,每一轮技术革新都引发了新一轮关于人类被工具替代的恐慌。技术日益深化的人工智能是否会导致大规模失业?从宏观理论分析来看,延续卡斯特的发现,人工智能对就业的影响呈现以下三种倾向。

1. 人工智能对就业的替代率更高,造成大规模失业

在一些特定领域,人工智能取代劳动力可能对社会产生积极影响,如降低商品价格并增加实际收入。然而,如果去技能化现象在所有领域同时发生,就业萎缩将对社会产生负面影响(Autor, 2015)。目前,机器人取代制造业劳动力的趋势已变得司空见惯。例如,拥有 120 万员工的富士康已用机器人取代了 6 万名人类员工(Wakefeld, 2016)。此外,到 2030 年,无人驾驶技术的普及将显著降低运输、采矿等部门中司机职业的重要性(Frey & Osbourne, 2017)。在美国,卡车运输业雇佣了 350 万人,是 29 个州最大的就业来源(Solon, 2016)。弗雷等人的研究表明,美国有 47%的工作属于高风险被替代类别,即未来 20 年内容易被自动化取代的工作,而另外 53%的工作(如医生、律师、工程师、教师等)属于低风险自动化类别(Frey & Osborne, 2017)。然而,这只是一个相对的分类,因为这些专业工作也可能面临自动化压力(Susskind & Susskind, 2015)。此外,弗雷等人的研究还表明,在这一波人工智能浪潮中,失业风险与平均工资中位数和教育水平呈显著负相关,即低技能工人与高技能工人之间的两极分化和不平等将进一步加剧,某些工人群体将不成比例地受到人工智能和自动化的影响。从企业

成本的角度来看,人工智能具有显著的成本优势。早在工业时代,马克思(2018:450)就在《资本论》中详细论证了机器替代人类劳动力的程度。例如,资本家仅需花费 1/4 先令即可使用蒸汽犁上的蒸汽机,这相当于花费 15 先令雇佣 66 个人劳动 1 小时。以智能客服领域为例,其在降低运营和人力成本方面的作用日益显著。数据显示,智能客服能够降低 30%—80%的运营成本(Greenwood, 2020)。预计到 2026 年,人工智能将帮助全球客服中心降低超过 800 亿美元的劳动力成本(朱江、孙蓟潍、郝毅杰等,2024)。

2. 人工智能创造就业

人工智能通过创新效应推动经济发展,具体表现为增加总体产出、释放对专业技术与服务的需求以及降低总体成本(Susskind & Susskind, 2015)。此外,在人工智能驱动的商业和工作中,将催生三种新型职业,威尔逊等人(Wilson, Daugherty & Morini-Bianzino, 2017)将其定义为培训师(trainers)、解释者(explainers)和维护者(sustainers)。培训师的主要任务是指导人工智能系统的运行,如训练客户服务聊天机器人识别人类交流中的复杂性和微妙之处,并表现出同情与理解。随着人工智能系统不透明性的增加,许多非技术人员对复杂机器学习算法的"黑箱"特性感到不安,这导致了解释者的出现。解释者的职责是弥合技术、商业、政府和社会之间的鸿沟,尤其是在系统建议的行动与传统价值观相冲突时。事实上,政府已开始考虑这一领域的监管。例如,欧盟于 2018 年生效的《通用数据保护条例》创造了一种"解释权",允许消费者质疑和反对完全基于算法的决策。维护者则负责确保人工智能系统按预期运行,预防意外后果,并确保其符合道德规范。基于 2010—2018 年的 Burning Glass 数据,奥托尔等人(Autor, Iazell & Restrepo, 2022)发现,2015 年后与人工智能相关的岗位空缺显著增加,这表明人工智能技术与企业需求的结合日益紧密;与此同时,与人工智能无关的职位明显减少,这一现象与人工智能替代部分劳动力技能的假设相一致。这

引出了第三个倾向,即就业替代与就业创造并存。

3. 替代和创造就业并存

为研究人工智能对就业的冲击,奥托尔等人通过对劳动力进行分类,探讨了其被人工智能替代的可能性。他们首先将劳动力划分为一个 2×2 的矩阵,其中 *x* 轴表示工作的认知程度,*y* 轴表示工作的重复程度(Autor, Levy & Murnane, 2003)。

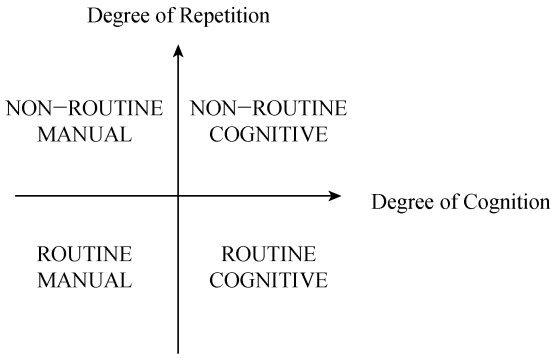


图 1 劳动与就业分类图

图片来源:Autor, Levy & Murnane, 2003.

由图 1 可知,常规/非常规类别指的是一项工作遵循明确规则的程度,而手工/认知分类则指工作对体力或知识能力的需求程度。*y* 轴以下代表了已经受到现代计算机技术冲击的工作,因为计算机在执行遵循预定义规则的任务方面比人类更具优势。例如,基础制造业属于常规体力劳动,而文字处理和计算则属于常规认知劳动。在这两类行业中,自动长途货车、自动客服和工业机器人将在短期内替代人力(Levy, 2018)。*y* 轴以上代表了需要高度直觉和问题解决能力的工作,这也是近年来发达国家就业增长的主要领域,如服务业和高技能职位(Autor, Levy & Murnane, 2003)、欺诈检测中的公正决策(Phua, Lee & Smith-Miles et al., 2012),以及律师助理和合

同/专利法研究系统(Markoff, 2011)。然而,福特(Ford, 2015)认为这种优势只是暂时的,人工智能和自动化技术将渗透到劳动力的所有领域,这种情况在历史上前所未有,并将对整个就业领域构成威胁。

为了回答“技术将替代就业还是创造就业”的问题,卡斯特关于“就业减少条件”的观点为我们提供了重要启示。他(卡斯特,2001:309)认为,在信息技术的传播下,只有两种情况下就业会减少:1. 需求的扩张未能跟上劳动生产力的增长;2. 制度未能针对这种不匹配的状况做出削减劳动时间的反应,而是选择削减工作岗位。

(1) 从 G7 国家与中国 2012—2022 年的数据(图 2)来看,国内总需求

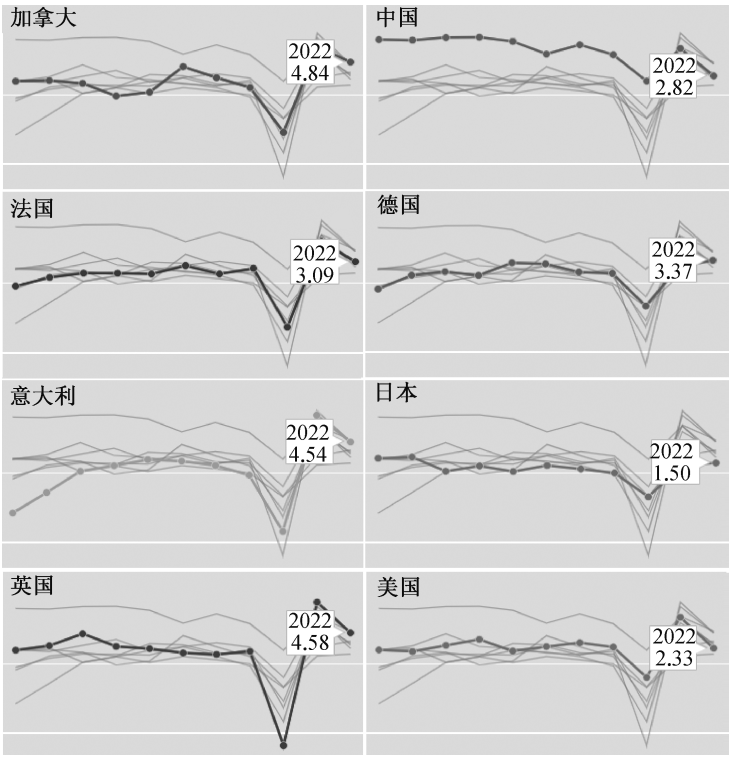


图 2 2012—2022 年 G7 国家国内总需求增长率(%)

图片来源:OECD, 2024a.

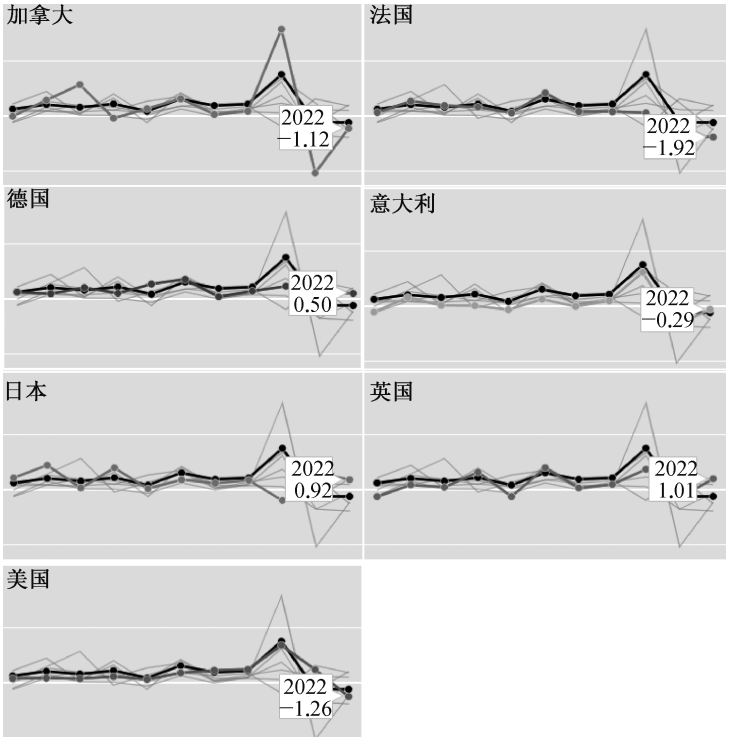


图 3　2012—2022 年劳动生产率年增长率(%)

图片来源:OECD, 2024b.

在 2019 年前总体呈增长趋势。受新冠疫情影响,所有国家的总需求在 2020 年降至负增长,但在 2022 年增长率回升至原有水平,甚至超过了 2018 年的增长水平。从劳动生产率(图 3)来看,除 2022 年受新冠疫情影响外,劳动生产率总体呈小幅稳定上升趋势,且其提升速度普遍低于总需求的增长速度。从失业率(图 4)来看,G7 国家的总体失业率逐年下降,尽管意大利、法国的失业率接近警戒线水平。

(2) 从劳动时间(图 5)来看,加拿大、德国、日本和意大利的工作时长稳步减少,而法国、英国和美国在过去 20 余年里工作时长变化不大。不过,

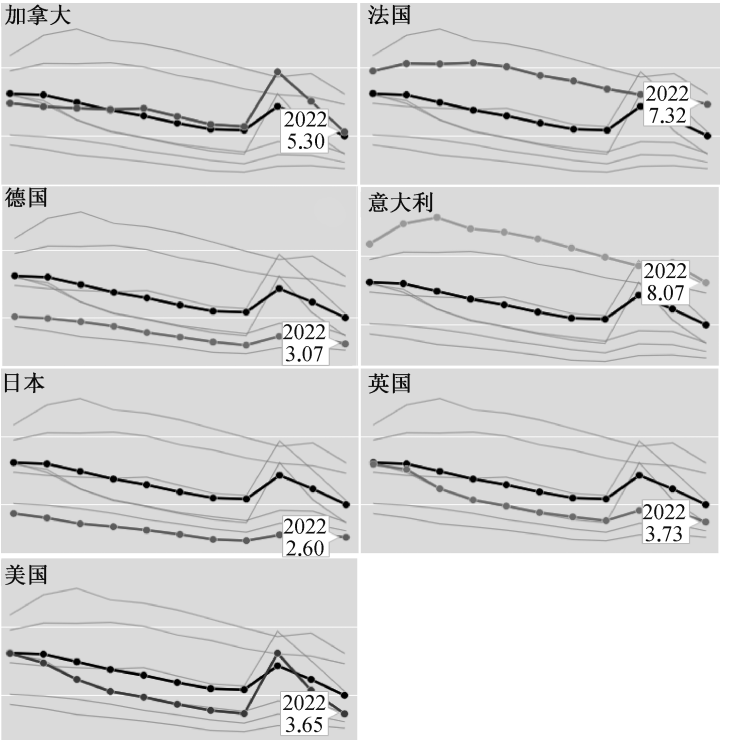


图 4　2012—2022 年 G7 国家失业率(%)

图片来源:OECD, 2024c.

英国、法国和德国的总体工作时间远低于 OECD 国家的平均水平,其中德国的工作时间还在逐年减少。

从宏观数据看,除受新冠疫情影响的几年之外,2012—2022 年 G7 国家社会总需求和劳动生产率总体呈上升趋势,且劳动生产率的增速并未超过总需求的增速;劳动时间逐年减少,失业率小幅下降。因此,卡斯特提出的就业减少的条件在近 10 年间并未在 G7 国家出现。从工作时间、工作效率和失业率三个统计数据来看,在数字经济高速发展的这些年里,并未造成大规模的失业。

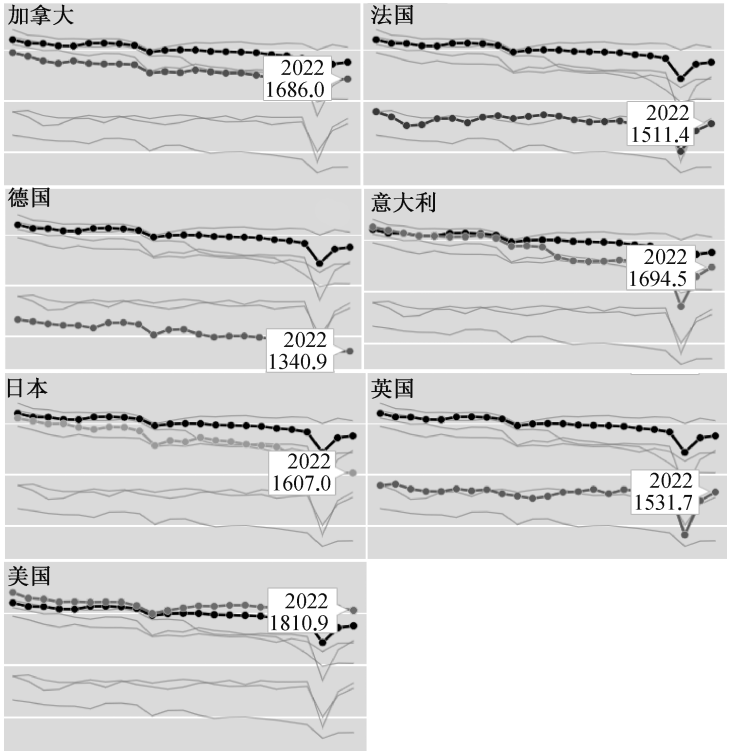


图 5　2000—2022 年 G7 国家年平均工作时间

图片来源：OECD，2024d.

五、总结与展望

本文立足于卡斯特关于网络社会的研究视域，分别探讨了在网络社会和智能社会中，技术与其他因素如何共同重构生产力和生产关系，以及如何重塑劳动和就业结构。

与工业社会相比，无论是 20 世纪 70 年代起步的信息科技，还是 2010 年后进入快车道发展的人工智能技术，它们作为新质生产力，先后被高新科

技巨头所掌握,并与其背后的支撑力量——主权国家共同构建起以它们为核心的、科技与资本高度集中的世界版图。而到了人工智能时代,这种垄断性特征将更加显著,权力结构也将变得更加牢固。

有关技术替代人类劳动力的焦虑,自工业革命以来便已存在。信息技术的扩散虽然取代了一部分劳动力并消除了某些工作岗位,但并未造成大规模失业。即使在数字经济蓬勃发展的今天,全球范围内仍未出现大规模失业的现象。尽管新冠疫情对社会造成了巨大冲击,但从数据来看,全球经济正在逐步恢复至疫情前的常态。在人工智能迅速发展的时代背景下,技术渗透到各行各业,影响就业结构和格局已成为不可逆转的趋势。

本文还存在诸多不足。首先,关于智能社会生产关系的论述,主要集中在生产资料的所有制形式及其引发的垄断性后果上。未来的研究方向可进一步关注以下几个方面:第一,人工智能对企业组织方式的影响;第二,通过解构劳动过程,探讨机器与人类劳动分工与合作的关系;第三,从工业时代的“劳动—资本”关系向智能时代的“机器和资本—劳动”关系的矛盾转变。此外,还可以探讨掌握新质生产力的群体是否会形成一种新的职业阶层,这种职业分层将如何形成,以及他们是否掌握上向社会流动的关键(布劳、邓肯,2021)。

其次,就劳动和就业结构的转变方式而言,本文仅提供了一个初步的视角。若要深入探究,还需细化更多问题,比如:如果人工智能技术替代了一部分就业岗位,那么这部分被替代的劳动力将流向何处?新的人工智能产业能够吸纳多少就业人口?这一动态过程需要多长时间才能完成?此外,人工智能可能带来哪些风险?

参考文献

贝尔,1984,《后工业社会的来临:对社会预测的一项探索》,高銛等译,北京:商务印书馆。

波兰尼,2020,《大转型:我们时代的政治与经济起源》,冯钢、刘阳译,北京:当代世界出版社。

布劳、邓肯,2021,《美国的职业结构》,李国武译,北京:商务印书馆。

陈刚、谢佩宏,2020,《信息社会还是数字社会》,《学术界》第 5 期。

国务院发展研究中心国际技术经济研究所,2019,《人工智能全球格局:未来趋势与中国位势》,北京:中国人民大学出版社。

吉登斯,2016,《社会的构成:结构化理论纲要》,李猛、李康译,北京:中国人民大学出版社。

卡斯特,2001,《网络社会的崛起》,夏铸九、王志弘译,北京:社会科学文献出版社。

李达,1958,《生产力与生产关系》,《理论战线》第 3 期。

马克思,2018,《资本论》,北京:人民出版社。

马克思、恩格斯,2016a,《马克思恩格斯全集》第 23 卷,北京:人民出版社。

马克思、恩格斯,2016b,《马克思恩格斯全集》第 46 卷,北京:人民出版社。

马克思,2018,《资本论》,北京:人民出版社。

邱泽奇,2022,《数字社会与计算社会学的演进》,《江苏社会科学》第 1 期。

王水兴、刘勇,2024,《智能生产力:一种新质生产力》,《当代经济研究》第 1 期。

卫兴华、李凯明,1980,《关于生产力的内容和发展生产力的问题》,《哲学研究》第 11 期。

杨善华,2016,《发展社会学通论》,北京:社会科学文献出版社。

赵家祥、聂锦芳、张立波,2003,《马克思主义哲学教程》,北京:北京大学出版社。

赵家祥,1989,《生产力系统和生产力发展动力》,《北京大学学报(哲学社会科学版)》第 1 期。

朱江、孙蓟潍、郝毅杰等,2024,《人工智能客服不“智能”如何解决?专家解读》,https://news.cyol.com/gb/articles/2024-03/15/content_AjX77ZUzLo.html。

Agrawal, A., J. Gans & A. Goldfarb 2019, “Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence.” in A. Agrawal, J. Gans & A. Goldfarb

eds. , *The Economics of Artificial Intelligence*: *An Agenda*, Chicago: University of Chicago Press.

Agrawal, A. , J. McHale & A. Oettl 2019, "Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth." in A. Agrawal, J. Gans & A. Goldfarb eds. , *The Economics of Artificial Intelligence*: *An Agenda*, Chicago: University of Chicago Press.

Allen, R. 2009, "The Industrial Revolution in Miniature: The Spinning Jenny in Britain, France, and India." *The Journal of Economic History* 69(4).

Aronowitz, S. & W. Di Fazio 1994, *The Jobless Future*, Minneapolis: University of Minnesota.

Autor, D. 2015, "Why are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29(3).

Autor, D. , F. Levy & R. Murnane 2003, "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118(4).

Autor, D. , J. Iazell & P. Restrepo 2022, "Artificial Intelligence and Jobs: Evidence from Online Vacancies." *Journal of Labor Economics* 40(S1).

Bartelsman, E. , M. Falk & E. Hagsten et al. 2019, "Productivity, Technological Innovations and Broadband Connectivity: Firm-Level Evidence for Ten European Countries." *Eurasian Business Review* 9.

Bessant, J. 1989, *Microelectronics and Change at Work*, Geneva: International Labour Organization.

Brynjolfsson, E. 2010, "Information Technology and the Reorganization of Work, Paper Delivered at a Conference on 'Vernetzung als Wettbewerbsfaktor'." https://www. academia. edu/2773577/Information_Technology_and_the_Reorganization_of_Work_Theory_and_Evidence.

Cockburn, I. , R. Henderson & S. Stern 2019, "The Impact of Artificial Intelligence on Innovation." in A. Agrawal, J. Gans & A. Goldfarb eds. , *The Economics of Artificial Intelligence*: *An Agenda*, Chicago: University of Chicago Press.

Commission of the European Communities 1994, *Growth, Competitiveness, Employment: the Challenges and Ways Forward into the 21st Century*, Luxembourg: Office of the European Communities.

Damioli, G., V. van Roy & D. Vertesy 2021, "The Impact of Artificial Intelligence on Labor Productivity." *Eurasian Business Review* 11(1).

Ford, M. 2015, *Rise of The Robots: Technology and The Threat of a Jobless Future*, New York: Basic Books.

Frey, C. & M. Osborne 2017, "The Future of Employment: How Susceptible are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114.

Goodman, P. & L. Sproull 1990, *Technology and Organization*, San Francisco: Jossey-Bass.

Gordon, R. 2016, *The Rise and Fall of American Growth: The US Standard of Living Since the Civil War*, Princeton: Princeton University Press.

Gordon, R. 2018, "Why Has Economic Growth Slowed When Innovation Appears to be Accelerating?" DOI:10.3386/w24554.

Greenspan, A. 1998, "The Semi-Annual Monetary Policy Report Before the Committee on Banking and Financial Services of the US House of Representatives."

Greenwood, A. 2020, "How AI Lowers Customer Service Costs." https://logic2020.com/insight/virtual-agent-customer-service-costs.

Gries, T. & W. Naudé 2018, "Artificial Intelligence, Jobs, Inequality and Productivity: Does Aggregate Demand Matter?" DOI:10.2139/ssrn.3301777.

Harrison, B. 1994, *Lean and Mean: The Changing Landscape of Corporate Power in the Age of Flexibility*, New York: Basic Books.

Imai, K. 1990, *The Information Network Society*, Tokyo: Chikuma Shobo.

Japan Institute of Labour 1985, *Technological Innovation and Industrial Relations*, Tokyo: JIL.

Jones, B. 2009, "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies* 76(1).

Kaplinsky, R. 1986, *Microelectronics and Work Revisited: A Review, Report Prepared for the International Labor Organization*, Brighton: University of Sussex Institute of Development Studies.

King, A. 1991, *The First Global Revolution: A Report by the Council of the Club of Rome*, New York: Pantheon Books.

Lee, K. 2016, "Artificial Intelligence, Automation, and the Economy." https://obamawhitehouse. archives. gov/blog/2016/12/20/artificial-intelligence-automation-and-economy.

Levy, F. 2018, "Computers and Populism: Artificial Intelligence, Jobs, and Politics in The Near Term." *Oxford Review of Economic Policy* 34(3).

Levy, R., M. Bowes & J. Jondrow 1984, "Technical Advance and Other Sources of Employment Change in Basic Industry." in F. Collins & L. Tanner eds., *American Jobs and the Changing Industrial Base*, Cambridge, MA.: Ballinger.

Markoff, J. 2011, "Armies of Expensive Lawyers, Replaced by Cheaper Software." http://www. nytimes. com/2011/03/05/science/o5legal. html.

Martin, C. 2000a, *Sustaining Flexibility: Work, Family and Community in the In-formation Age*, Cambridge, MA.: Harvard University Press.

Martin, C. 2000b, *Sustaining the New Economy: Work, Family, and Community in the Information Age*, Cambridge, MA.: Harvard University Press.

Marx, L. & M. Smith 1991, *Does Technology Drive History? The Dilemma of Technological Determinism*, Cambridge, MA.: MIT Press.

Miller, S. 1989, *Impacts of Industrial Robotics: Potential Effects of Labor and Costs within the Metalworking Industries*, Madison: University of Wisconsin Press.

Narechania, T. 2021, "Machine Learning as Natural Monopoly." https://ilr. law. uiowa. edu/sites/ilr. law. uiowa. edu/files/2023-02/A4_Narechania. pdf.

Nelson, R. 1994, "An Agenda for Formal Growth Theory." https://core. ac. uk/download/pdf/33895596. pdf.

OECD 1994, *Employment/Unemployment Study*, Paris: OECD.

OECD 2024a, "Domestic Demand Forecast." https://data.oecd.org/gdp/domestic-demand-forecast.htm.

OECD 2024b, "Labour Productivity and Utilisation (Indicator)." https://data.oecd.org/lprdty/labour-productivity-and-utilisation.htm.

OECD 2024c, "Unemployment." https://data.oecd.org/unemp/unemployment-rate.htm#indicator-chart.

OECD 2024d, "Hours Worked." https://data.oecd.org/emp/hours-worked.htm.

Phua, C., V. Lee & K. Smith-Miles et al. 2012, "A Comprehensive Survey of Data Mining-Based Fraud Detection Research." *Computers in Human Behaviour* 28(3).

Rheingold, H. 1993, *The Virtual Community: Homesteading on the Electronic Frontier*, Reading: Addison-Wesley.

Richard, F. 1994, *Working under Different Rules*, Cambridge, MA.: Harvard University Press.

Schettkat, R. & M. Wagner 1990, *Technological Change and Employment lnnovation in the German Economy*, Berlin: Walter De Gruyter.

Schiller, D. 1999, *Digital Capitalism: Networking in the Global Market System*, Cambridge, MA.: MIT Press.

Solon, O. 2016, "Self-Driving Trucks: What's the Future for America's 3.5 Million Truckers." https:/www.theguardian.com/technology/2016/jun/7/self-driving-trucks-impact-on-drivers-jobs-us.

Susskind, R. & D. Susskind 2015, *The Future of the Professions: How Technology Will Transform the Work of Human Experts*, New York: Oxford University Press.

Uchitelle, L. 1999, "Big Increases in Productivity by Workers." *The New York Timmes* November 13.

van den Besselaar, P. 1996, "Technology, the Future of Our Jobs and Other Justified Concerns." https://dl.acm.org/doi/pdf/10.1145/238339.238357.

Wakefeld, J. 2016, "Foxconn Replaces 60 000 Factory Workers with Robots." http://hwww.bbc.com/news/technology-36376966.

Wilson, H., P. Daugherty & N. Morini-Bianzino 2017, "The Jobs That Artificial Intelligence Will Create." https://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates.

Zysman, J. & L. Tyson 1983, *American Industry in International Competition: Government Policies and Corporate Strategies*, Ithaca: Cornell University Press.

智能社会研究
Journal of Intelligent Society

编 委 会 主 任：高　岩

编委会副主任：夏桂华　赵玉新
吕　鹏（中国社会科学院）

编　　委：尹　航　冯仕政　冯全普
（按姓氏笔画排序）　吕　鹏（中南大学）　吕冬诗
朱齐丹　汝　鹏　苏　竣
李正风　来有为　肖黎明
邱泽奇　何晓斌　宋士吉
陈云松　陈华珊　郑　莉
孟小峰　孟天广　赵万里
赵延东　胡安宁　袁　岳
黄　萃　梁玉成　董　波
曾志刚　蔡成涛

青 年 编 委：丁奎元　王　磊　叶瀚璋
（按姓氏笔画排序）　邢麟舟　向　维　刘灿辉
刘松吟　刘春成　刘晓波
安　博　许馨月　孙宇凡
李子信　李天朗　李晓天
吴雨晴　何　丽　邹冠男
张咏雪　张承蒙　陈　苗
陈典涵　林子皓　周雪健
周骥腾　郑　李　胡万亨
茹文俊　贺久恒　贾雨心
郭媛媛　黄　可　梁　轩
曾　晨

编辑团队

主　　编：郑　莉
编辑部主任：吴肃然
编辑部成员：林召霞　王立秋
李昕茹　李天朗
岳　凤

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：2024 年 11 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

本刊已许可中国知网等网络知识服务平台以数字化方式复制、汇编、发行、信息网络传播本刊全文。本刊支付的稿酬已包含网络知识服务平台的著作权使用费，所有署名作者向本刊提交文章发表之行为视为同意上述声明。如有异议，请在投稿时说明，本刊将按作者说明处理。