智能社会研究

（双月刊）

2024 年
第 1 期
2024 年 1 月 10 日出版

2022 年 11 月 10 日创刊

总第 8 期

目　次

论文

平台企业的灵活积累机制

——一个合约理论的视角 …… 滕　飞（ 1 ）

信息逆差：代码权力膨胀与公众权力让渡 …… 施颖婕（ 21 ）

大语言模型能用来给主题模型中的主题编码吗？

——兼论大语言模型在社会科学领域应用的前景

…… 李　代　张博伦　周沁莽（ 33 ）

数字时代的科技公司与国家治理

——一个综合性的分析框架 …… 钟瑞雪（ 60 ）

新闻大数据视角下测度区域一体化水平的新路径

——以长三角地区为例 …… 李　军　董方杰（ 84 ）

解密数字社会中的幸福密码

——不同互联网使用程度下社会信任感和社会公平感对主观幸福感的

影响机制研究 …… 郭媛媛　张　腾（103）

研究报告

数字时代安全科技价值报告 …… 陆菲菲　秦　昊　张友红　吕　鹏(133)

问题触发的算法模型响应机制探索 …… 许正军　袁　岳(153)

译文

数字健康的社会技术伦理

——生命伦理学进路的批判和延伸

…… 詹姆斯·肖　约瑟夫·多尼娅　著　何　丽　译(166)

书评

数字资本主义和智能技术的当代批判

——评《过度智能》 …… 周银知(186)

访谈

通过人类行动思考机器智能

——专访人类学家露西·萨奇曼 …… 露西·萨奇曼　胡万亨(198)

CONTENTS

THESIS

The Flexible Accumulation Mechanism of Platform Enterprises: A Contract Theory Perspective ····· Teng Fei(1)

"Information Deficit": Code Power Inflation and Public Power Cession ··· Shi Yingjie(21)

Can LLM Label Topics from A Topic Model: The Future of LLM in Social Science ····· Li Dai, Zhang Bolun, Zhou Yimang(33)

Technology Companies and National Governance in the Digital Age: A Comprehensive Analytical Framework ····· Zhong Ruixue(60)

A New Approach to Measuring Regional Integration from the Perspective of Big Data in News: A Case Study of the Yangtze River Delta Region ··· Li Jun, Dong Fangjie(84)

Decrypting the Happiness Code in the Digital Society: A Study on the Influence Mechanism of Social Trust and Fairness Perception on Subjective Well-Being under Different Internet Usage Levels ····· Guo Yuanyuan, Zhang Teng(103)

RESEARCH REPORTS

Report on the Value of Security Technology in the Digital Age ····· Lu Feifei, Qin Hao, Zhang Youhong, Lyu Peng(133)

Exploration of Algorithmic Model Response Mechanisms for Problem Triggering ····· Xu Zhengjun, Yuan Yue(153)

TRANSLATED TEXT

The Sociotechnical Ethics of Digital Health: A Critique and Extension of Approaches from Bioethics ····· written by J. Shaw, J. Donia; trans. by He Li (166)

BOOK REVIEW

A Contemporary Critique of Digital Capitalism and Smart Technologies: Review of *Too Smart* ····· Zhou Yinzhi (186)

INTERVIEW

Thinking Machine Intelligence Through Human Actions: An Interview with Anthropologist L. Suchman ····· L. Suchman, Hu Wanheng (198)

平台企业的灵活积累机制*

——一个合约理论的视角

滕　飞**

摘要：本文从新制度经济学合约理论的视角解释了平台企业的灵活积累机制，这一机制的运作基础是平台企业与大量外部专业化组织及零工群体建立的不平等合约关系。该机制的运作过程分为两个方面：其一，在不平等的合约关系下，平台企业得以将规模化扩张带来的大部分固定资产成本、人力成本和管理成本转移至外部专业化组织及零工群体，从而实现轻资产化运营；其二，平台企业为规避合约关系中的“委托—代理”风险，会综合运用平台规则、业务考核、大数据和算法等手段对外部专业化组织及零工群体构筑严密的管控体系，并严格规定产品/服务的生产过程和结果。由于合约关系的不完全性和合约关系中剩余控制权的存在，在平台企业的灵活积累机制之外，零工群体也会积极采取各种措施维护自身利益，构建自身主体性。

关键词：平台企业　合约理论　灵活积累　零工群体

近年来，平台企业在全球迅速发展。相关数据显示，2020 年全球市值超 100 亿美元的平台企业数量达 76 家，市值总额为 12.5 万亿美元，市值总额同比增长 57%；中美市值百亿美元以上的平台企业数量合计达 64 家，全球占比 84.2%（中国信息通信研究院，2021）。平台企业之所以具有极强的市场竞争优势并能够快速实现规模化扩张，一个关键原因就在于其生产组织方式和资本积累方式具有低成本投入、低风险承担的特征。本文指出，这种生产组织方式和资本积累方式实质上是一种后福特主义式的灵活积累机制。

这种灵活积累机制得以实现的基础是，平台企业综合运用市场合约、平台规

* 本文系中国社会科学院学科建设“登峰战略”资助计划新兴学科与交叉学科“科技与数字社会”的阶段性研究成果。感谢匿名审稿专家和编辑老师的宝贵修改建议，文责自负。

** 滕飞，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院。

则、业务考核、大数据和算法等手段，对与之合作的外部专业化组织[①]和零工群体实现严密的生产管控，平台企业还将规模化扩张所带来的大部分成本及风险转移至这些外部专业化组织和零工群体中。因此，平台企业在全球高速扩张的同时，所引发的社会后果却是让人喜忧参半的（吕鹏、周旅军、范晓光，2022），我们需要对平台企业的灵活积累机制做出合理分析，从而更好地应对平台可能给社会带来的负面影响。本文从新制度经济学合约理论的视角出发，在理论上解释了这种灵活积累机制的具体运作过程，这一理论解释适用于不同类型的数字平台，如内容平台、电子商务平台、服务性劳动平台[②]等。

一、平台企业后福特主义式灵活积累的发展

从工业革命到信息革命，发达国家企业的生产组织方式和资本积累方式经历了从福特主义（Fordism）向“后福特主义”式灵活积累（flexible accumulation）的转变。根据杰索普（B. Jessop）对相关文献的总结，福特主义作为一种典型的产品生产与资本积累模式，在劳动过程（the labor process）、积累机制（the regime of accumulation）、管制模式（modes of regulation）、社会化模式（modes of societalization）四个方面有着鲜明特征：劳动过程是基于泰勒制的专业化技术分工；积累机制是基于大规模生产和大规模消费不断增长的良性循环；管制模式是规范、制度、组织形式、社会网络和行为模式的集合，管制模式维持和指导福特主义的积累机制；社会化模式主要是指福特主义的社会影响，如促进了家庭对大规模标准化生产商品的消费、促进了工业区域的形成等（Jessop，2005）。

后来，福特主义因为一些因素逐渐退出历史舞台，包括生产效率下降、工人工资上涨、大宗商品国际市场饱和、产能的结构性过剩以及消费者多元化需求的

① 本文提及的“外部专业化组织”是指在组织架构上不隶属于平台企业，但是和平台企业进行专业化分工协作的组织。在各类数字平台的生态圈中，平台企业主要专注于数字平台基础设施的建设、平台治理规则的制定、用户流量的增长，而外部专业化组织则主要协助平台进行专业内容/服务的生产和供给，并进行流量的商业化变现。例如，在直播平台上，MCN 机构/公会就是平台企业的外部专业化组织，它们与平台合作培养主播，并根据平台规则进行专业内容的生产和用户流量的变现。

② 本文提及的“服务性劳动平台”是指那些提供各类服务性劳动的平台，如提供物品收派服务的物流快递平台、提供外卖取送服务的外卖平台、提供用车服务的网约车平台等。

增长、第三世界廉价劳动力的竞争、生产组织的官僚化、新的技术革命与产业发展等。由此，福特主义开始向后福特主义转型（Roobeek，1987）。

与福特主义相比，后福特主义式的企业在产品生产和资本积累上的典型特征是灵活积累。沃特金斯（P. Watkins）指出，这种灵活积累的灵活性体现在三个方面。首先是生产的灵活性。知识和信息在企业的生产环节发挥关键作用，企业将快速收集的信息与一个灵活的生产系统相结合，以便对市场变化做出快速反应；企业逐渐放弃垂直一体化的生产组织方式，选择与外部不同生产单元的专业化合作来减少生产成本，最终形成一个松散的、适应市场快速变化的生产联盟。其次是技术的灵活性。计算机和机器人等信息技术的广泛应用，允许企业将一体化的生产环节分解为许多独立的生产单元，这样还可以有效地管理这些分散的单元；基于信息技术的柔性控制也解放了管理层的双手，并实现了对劳动力更加严密的控制。最后是劳动用工的灵活性。企业普遍采用弹性雇佣策略，兼职工和临时工被公司大量使用。当工作量发生波动时，公司可以随时要求这些非正式员工提供相应的劳动服务，但公司却不用为这些非正式员工承担病假、年假、社会保障等福利开支（Watkins，1991：13-36）。

灵活的生产组织方式和资本积累方式给企业带来了新的竞争优势，使得企业可以对市场需求变化和市场竞争威胁采取快速灵活的反应，也使得企业可以在短期内灵活调整产品数量和样式（Schoenberger，1988）。在数字经济时代，得益于数字技术的赋能，平台企业进一步发展了后福特主义式的灵活积累，这体现为其灵活的生产组织方式、灵活的劳动用工以及基于此的低成本投入、低风险承担的规模化扩张与资本积累。

在生产组织方面，平台企业同样选择与外部专业化组织合作生产，而非采取垂直一体化的生产方式。但是平台企业与外部专业化组织的关系并非后福特主义式的松散联盟，而是在数字技术赋能下的紧密联盟。这种联盟关系以平台企业为主导，平台企业积极培育各类外部组织帮助其实现各类产品/服务的专业化生产和供给，并将不确定的经营风险和经营成本转移至这些外部组织，还通过市场合约、平台规则、业务考核、大数据和算法等手段对这些外部组织进行严密管控（黄贺铂，2021；叶韦明、金一丹，2021；赵磊、韩玥，2021；翟秀凤，2021；刘亭亭、杨晓兵，2022）。

在劳动用工方面，平台企业仍然采取后福特主义式的灵活用工策略，且在数字技术的加持下对劳动过程的控制更加严密（陈龙，2020a；梁萌，2021）。这些控制手段包括算法控制（陈龙，2020b）、时间控制（李胜蓝、江立华，2020）、游戏化控制（邓智平，2021）、意识形态控制（刘战伟、李媛媛、刘蒙之，2021；刘战伟、刘蒙之、李媛媛，2022）、量化考核控制（徐林枫、张恒宇，2019）等，导致了零工群体出现情感劳动（涂永前、熊赟，2019）、数据劳动（任桐、姚建华，2022）、商品化和工具化（蒋淑媛、黄彬，2020）、职业认同感低（刘蒙之、刘战伟，2021）等劳动异化现象。

二、合约理论视角下平台企业灵活积累的运作机制

认识到平台企业灵活积累的发展特征后，我们需要进一步对这种灵活积累的运作机制进行剖析。具体来说：平台如何与各类外部专业化组织以及零工群体形成合作关系？这种合作关系的本质是什么？在这种合作关系下，平台企业何以能够在轻资产、低成本投入、低风险承担的条件下快速实现规模化扩张和资本积累？下面，我们运用新制度经济学合约理论来分析平台企业灵活积累的运作机制。

在合约理论看来，任何交易关系都必须通过某种形式的合约来调解。“委托—代理”关系是合约关系的一种类型，这种关系是指由于市场交易中比较优势和专业化分工的存在，常由代理人代表委托人行事（Hart & Holmström，1987）。在“委托—代理”关系中，代理人必须从多种可能的行动中选择一个，该行动将影响代理人及其委托人的利益；委托人具有制定支付规则的权利，这一支付规则是在代理人采取行动前，委托人根据对代理人行为的观察和预测制定的。因为信息和知识不对称、目标和利益不一致，“委托—代理”关系中又常发生道德风险（代理人借事后信息不对称、不确定以及合约的不完全性而采取不利于委托人的行为）和逆向选择（代理人利用事前信息不对称所进行的不利于委托人的决策选择）（Arrow，1984）。

平台企业实质上是一个基于多边合约关系的网络化组织。平台企业的运营模式与传统企业的不同之处在于：平台企业是价值的整合者、多边群体的连接

者,平台商业模式的根基来自多边群体的互补需求所激发的网络效应(陈威如、余卓轩,2013:82),具体来说,其中一类群体参与市场交易所获得的收益会受其他几类群体数量或质量的影响,平台上的消费者增多会吸引更多的生产者加入以获取更多商机;反之,更多的生产者加入以及服务或商品的多样化会吸引更多的消费者加入(帕克、埃尔斯泰恩、邱达利,2017:29—31)。在网络效应驱动下,平台方对数据进行搜集、归类、分析、计算与预测等,使之服务于更为高效、精准的资源配置与信息传递,同时使数据要素反哺平台发展,进一步建设、维护和改进数字平台的规则和基础设施,以吸引更多用户(王勇、戎珂,2018:2—3;张雪娇、肖潇,2021)。

但是,平台用户规模突破临界点是网络效应形成正反馈的前提。当用户规模没有达到临界点时,市场会逐渐萎缩甚至为零;但只要用户规模超过临界点,便会激活正反馈机制,正反馈又会产生消费者的转移成本和锁定效应,产生赢家通吃的市场现象(傅瑜,2013:10)。一般来说,平台用户规模越大,平台在基础设施建设、人力、物力上的投入也会越大。因此,在扩大用户规模的同时控制平台建设成本,是所有平台在扩张时必须解决的问题。

平台一般通过与外部专业化组织、零工群体建立不平等的合约关系来转移规模化扩张带来的成本与风险,并实现轻资产运营的目的。从生产资料的占有和使用来看,平台通过与大量的平台用户建立合约关系,实现对平台用户生产资料的整合,这大大降低了平台在扩张中的不确定风险和成本投入。平台掌握着流量、大数据、匹配技术等优势生产资料,并具有对生产资料的集中与整合优势,这使得平台用户不得不依附于平台并共享自身生产资料的使用权(邵占鹏,2022)。在平台用工方面,数字平台上的劳动力供给一般采取“平台+组织+个人”的三角式用工模式,平台通过人力资源外包来实现低成本、低风险的灵活用工(王茜,2020;Kuhn & Maleki, 2017)。因此,平台企业通过与大量外部主体建立不平等的合约关系,使得其固定资产成本、人力成本、管理成本[①]等成本投入比传统企业少得多,平台得以在轻资产、低成本投入、低风险承担的条件下快速

① 在新制度经济学看来,管理成本包括两个方面:一是设立、维持或改变组织设计的费用,即管理制度交易费用,这种费用与一系列活动有关;二是组织运行的费用,即管理制度交易费用给定下具体管理交易的费用(袁庆明,2021:44)。

实现规模化扩张和资本积累。

但是,合约关系也不可避免地具有“委托—代理”关系的各种风险(如前所述的逆向选择、道德风险等)。为了规避“委托—代理”关系中的风险,平台企业会综合运用平台规则、业务考核、大数据和算法等手段对外部专业化组织及零工群体进行管控。

以网约车平台为例,平台选择通过市场方式将数字技术无法完成的非标准化司机管理职能交由租赁公司完成,这既能直接减少用工成本,又能降低司机招募和管理过程中的交易成本(如搜寻成本、协调成本等),还能阻断因直接管理司机而被认定为雇佣关系的法律风险和用工成本;再通过控制和激励手段(包括行为规则、绩效考核和奖惩、培训以及数字技术等)对“委托—代理”关系下可能产生的逆向选择、道德风险进行规避,最终构筑了由“平台企业—租赁公司—队长—网约车司机”组成的层级管理结构(赵磊、韩玥,2021)。

再以家政平台为例,平台为家政工提供市场需求信息,其与劳动者是一种弱契约关系,但平台在实际运营中却借助平台规则及消费者评价对劳动者实施与权利关系不对等的强控制,平台企业构建了强控制弱契约的用工模式(梁萌,2017)。在直播平台上,平台通过专职运营人员以“返点政策”来控制直播公会的日常运营,并借助算法技术对直播公会的日常运营进行流量奖励或处罚;而直播公会则会将平台的控制压力转移给主播,通过薪酬激励、流量扶持、合约限定以及情感支持等方式来加强对主播内容生产的控制(刘倩,2023)。

在劳动过程理论看来,企业对劳动者的控制与劳动者的反抗一直是相伴而行的(孙兆阳,2013)。而在新制度经济学看来,劳动者的反抗或者自由行动空间是由合约关系中的剩余控制权带来的。合约关系的不完全性,必然导致“委托—代理”关系中代理角色拥有剩余控制权(Hart, 2017)。剩余控制权是指那种事前没有在合约中明确界定如何使用的权利,即资产所有者可以按任何不与先前的合同、习惯或法律相违背的方式决定资产用法的权利(傅绍文、邓秋云,2004)。

因此,无论是从劳动过程理论来看,还是从新制度经济学的合约理论来看,平台上作为代理角色的零工群体都是具有一定剩余控制权和自由行动空间的。尽管平台企业借助市场合约、平台规则、业务考核、大数据和算法实现了控制手

段的升级，但仍然无法对劳动者所处劳动场景中的变化进行完全预测和控制，劳动者会充分调动自主性为自己争取更多的自由行动空间（张杨波，2020；付堉琪，2021），平台对劳动过程的控制和劳动者拥有工作自主权是并存的（吴清军、李贞，2018）。因此，处于代理角色的外部专业化组织及零工群体会在有限的空间内与平台展开博弈以争取自身利益。为了更直观地理解平台企业灵活积累的上述运作机制，我们将上述逻辑用图 1 来展示。

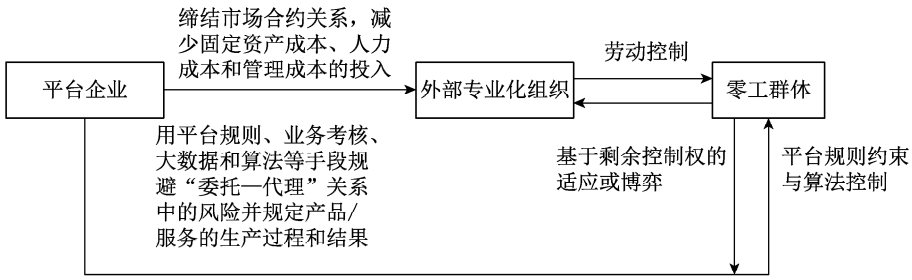


图 1　合约理论视角下平台企业灵活积累的运作机制

三、典型平台企业的灵活积累

在内容平台、电子商务平台以及各类服务性劳动平台上，平台企业灵活积累的运作机制皆有体现。无论平台对用户的主要供给品是内容、商品还是服务性劳动，平台首先都会积极培育、扶持各类外部专业化组织，与之建立不平等的市场合约关系，并将平台规模化扩张所带来的大部分固定资产成本、人力成本、管理成本负担转移至这些外部专业化组织，平台得以在轻资产、低风险的条件下实现规模化扩张与资本积累。此外，平台广泛运用平台规则、业务考核、大数据和算法等手段对外部专业化组织的运营实践进行管控，并要求这些外部专业化组织按照平台要求对零工群体的劳动过程进行管控。同时，平台还通过平台规则和算法对零工群体的劳动过程实施隐蔽的监视和管控。

（一）内容平台

以抖音、快手为代表的内容平台，积极培育 MCN 机构/公会等外部专业化

组织来帮助平台实现内容生产的专业化并协助平台对零工群体实施生产规训，“平台+MCN 机构/公会+零工群体”的多级内容生产和管控体系得以建立。以内容平台中的直播业务为例。首先，直播公会作为重要的再中介者，通过规范化管理和专业化培养主播、深耕直播细分市场并利用平台界面创生娱乐新业态，起到了降低平台成本、增加平台收益的重要作用（刘亭亭、杨晓兵，2022）。其次，公会还起到了最为关键的生产组织和劳动协调作用。公会里的经纪人和运营人员通过对内的生产组织（招募、培训和管理）和对外的可见性管理（运营、导流和打榜），向平台提供一定数量和质量的内容，帮助平台实现专业化、职业化、商品化的直播产业塑造（叶韦明、金一丹，2021）。

由于资源、信息和权力地位的不对等，MCN 机构/公会的日常运营是在平台的严密管控下进行的。平台在商业议价、政策制定与审查、控制内容制作方向等方面具备绝对权力，平台建构的加速时间秩序已经深度影响了 MCN 机构/公会的运作模式和行业实践（黄贺铂，2021）。在平台的创新机制下，MCN 机构/公会常常被动跟随平台进行规则变换和战略调整。借助算法技术和资本运作，平台能够迅速建立并推广 MCN 机构/公会的变现机制，现有机制的盈利效率一旦下降，平台就可以通过算法技术和资本运作迅速废弃现有机制并建立新的机制（翟秀凤，2021）。

此外，平台还借助 MCN 机构/公会这些专业化组织实现内容生产所需的弹性化劳动力供给，作为劳动后备军的零工群体处于极不稳定的状态中。在各类内容平台中，作为职业化内容生产者的网红圈层呈现金字塔的结构，头部网红获得平台的流量倾斜和主要孵化资源，尾部网红则面临商业模式不匹配、人设崩塌、上升空间狭窄等问题（蒋文妍，2020），大部分网红群体存在身份焦虑和社会认同感缺失等问题（常珂璨，2018）。在平台和资本的裹挟下，内容生产者逐渐陷入对内容生产量的追求，并自我固化了其内容输出的模式，这导致其陷入有增长而无发展的内卷化困境（牛天，2022）。

（二）电子商务平台

与内容平台的“平台+MCN 机构/公会+零工群体”多级内容生产和管控体系类似，电子商务平台（以下简称“电商平台”）形成了“平台+中小企业/工厂+

灵工群体”的商品生产和供给模式。近年来，用户直连制造（customer to manufacture，C2M）作为消费互联网和产业互联网融合的新型生产制造模式开始兴起。在该模式下，电商平台或直播机构基于用户需求的实时反馈或大数据分析来规划上游生产活动，以需定产逐渐取代以供定产的业务分包模式，上游厂商的生产制造模式和劳动关系在消费者需求的驱动下得到重塑。

虽然C2M模式降低了商品流通领域的成本并使得商品更加符合消费者需求（刘晓丽、王志勇，2020），但是由于消费者需求的不确定性，电商平台的上游厂商、生产作坊和劳动者实际上承担着灵活多变的订单式生产带来的风险与成本，这是电商平台利用市场合约关系转移生产不确定性的典型表现。以服装行业为例，为了满足电商平台在生产方面提出的低加工费、小批量、多品种、快返单的要求，产业链上形成了“大工厂+小工厂+作坊”的柔性化协作模式，劳动关系呈现出“正规+非正规+极致的灵活”多种形式，这是一种平台占主导地位的新型柔性生产组织方式（范璐璐、黄岩，2021）。

此外，就平台与各类网商的关系而言，表面上平台通过提供线上渠道与网络零售商、商品生产商形成平等的市场合约关系，但平台掌握着流量、大数据、匹配技术等优势生产资料，因此具有多种生产资料的集中与整合优势；网商与其生产商品、运营网店的生产资料不得不依附于平台，尤其是依附于处于寡头或垄断位置的大型平台（邵占鹏，2022）。为了保证平台的声誉以及稳定的商品供给，平台对网商群体实施严密管控。作为线上交易平台的电商平台兼具市场属性，因此也具有信息不对称、机会主义、逆向选择和道德风险等市场失灵现象，平台运营主体则通过制定市场规则，采取信号显示机制、声誉机制、质保机制和第三方介入机制来确保平台交易的健康有序（汪旭晖、张其林，2017）。平台型电商企业与平台卖家建立了介于科层制与市场制间的中间型关系，平台型电商企业在社会大市场下自主设计平台小市场，规范平台卖家行为，引导平台卖家成长（汪旭晖、张其林，2016）。

（三）服务性劳动平台

内容平台提供各类在线内容的生产与供给服务，电商平台提供商品交易服务，本文所指的各类服务性劳动平台（如物流快递平台、外卖平台、网约车平台）

则提供灵活的服务性劳动(如物品收派、外卖取送、用车服务)。服务性劳动平台所处行业为劳动密集型行业,平台为了减少规模化扩张所带来的固定资产投入、人力成本以及管理成本,积极与各类外部专业化组织建立不平等的合约关系,并通过平台规则、业务考核、大数据和算法等手段对这些外部组织及零工群体进行严密控制。

与电商业发展联系紧密的物流快递平台采取加盟模式快速扩张,其组织形态呈现出以总部为核心、以加盟商为延伸的网状结构,灵活专业的组织间关系为总部带来了极大的好处:在组织管理方面,在电子面单和菜鸟指数评价体系等数字技术的支撑下,总部得以对快递实现全链路的品质管理,降低了因组织规模过大而带来的管理成本;在成本转移方面,加盟商为总部节省了网点、仓储等固定资产以及人力成本,加盟费却给总部带来了巨大的外部融资收入,总部还对加盟商执行绩效考核和淘汰制度来保障加盟商的运营质量;在应对市场变化带来的风险方面,加盟商享有的部分柔性定价权解决了总部中心定价难的问题,加盟商内部的熟人式管理和自负盈亏模式也保证了高强度、高弹性的劳动供给(沈晓峰、袁钉,2020)。

与物流快递平台类似,以某网约车平台为例,该平台凭借垄断地位在企业外部构建了一个由多市场主体组成的、层级化的科层控制体系,平台先在企业外部设置岗位并通过合作方式组织服务的供给,通过合约规则确保网约车司机和租赁公司行为的可预见性,通过数字技术和租赁公司承担司机管理职能并执行平台规则(赵磊、韩玥,2021)。再以某外卖平台为例,该平台通过建立大规模网络化的配送站点来保证充足而灵活的劳动力供给,平台会对外卖配送站点实施量化考核以及末位淘汰制,以督促站点对骑手的劳动过程进行量化管理(赖祐萱,2020)。

此外,为了降低劳动力成本,外卖平台还与第三方用工机构(third-party staff agencies, TSAs)合作,将劳动力招聘和管理外包给 TSAs。TSAs 负责管理某些地区的骑手并采取自上而下的等级式管理,这个等级式管理体系由管理中心、区域、商圈、站点和小组组成;TSAs 还协助平台在线上实时进行劳动管理活动,包括背景和健康检查、在线培训、信息传播和管理会议等。在 TSAs 的帮助下,外卖平台实现了业务的低成本、规模化扩张(Sun, Chen & Rani, 2023)。在不同的

外卖平台中，平台企业会综合借助加盟承包商、灵活用工服务商、保险公司等外部组织实现外包式用工（李强、徐兴荣、刘凤杰，2022）。

除了借助各类外部专业化组织对劳动者进行管控外，平台更多地借助平台规则、大数据和算法对劳动者的任务分配、劳动过程、绩效考核和成果分配实施隐蔽化、精细化的管控。算法控制是在线劳动力平台运营的核心（Wood，Graham & Lehdonvirta et al.，2019），平台基于平台数据的算法自动管理与用户评分和评价机制，实现对劳动者低成本、高效率的监督管理，并通过引导劳动者提高劳动强度来促进自身的资本积累（谢富胜、吴越，2019）。这充分解决了平台因规模庞大而导致的管理难题，大数据和算法作为“无形的手”在进行管理。

具体来说，平台首先建立起一套量化的、算法驱动的劳动收入体系。以某外卖平台为例，该平台将接单量、超时率、差评率、投诉率作为站点及骑手的关键绩效考核指标：跑的单越多，准时率越高，顾客评价越好，骑手获得的积分便会越高；积分越高，等级就越高，奖励收入也会更多（赖祐萱，2020）。因此，平台通过制定差序格局和游戏化的管理方式，使得劳动者陷入一种自我剥削的循环（孙萍，2019），并将平台与劳动者之间的冲突转化为劳动者之间的游戏化竞争（杨善奇、刘岩，2021）。此外，平台通过消费评价机制发动平台用户对劳动者服务质量和劳动过程进行监控，劳动者在劳动过程中受到来自企业和用户的双重监管，管理控制程度被大大加强（梁萌，2017）。通过消费评价机制，平台与劳动者之间的冲突转化为劳动者与消费者之间的冲突（王蔚，2021）。

平台还通过算法的动态定价来调动劳动力的动态供给。平台作为一个多边市场，不仅需要考虑向一类顾客收取的价格水平，还需要考虑向另外一类顾客收取的价格水平，以及两类顾客价格水平的相对变化对平台交易规模的影响（王勇、邓涵中，2017）。各类零工平台通过动态定价的方式，实时调节顾客需求和劳动力供给，以有效保证整个系统的稳定性与用户的良好体验（刘善仕、裴嘉良、葛淳棉等，2022；陈龙、韩玥，2020）。在平台算法的管理下，平台上的零工劳动者成了平台和算法的机械延伸，成了平台和算法所管理的一种招之即来挥之即去的客观要素（谢富胜、吴越，2019）。

（四）小结

根据已有研究成果，我们发现构建灵活积累机制已经成为平台规模化扩张的必选路径。平台企业与各类外部组织及零工群体建立外包/业务合作的合约关系，有利于平台降低固定资产成本、人力成本和管理成本，也有利于平台根据市场环境变化和平台战略变化及时调整对外部组织及零工群体的要求，而变化和调整带来的不确定风险及成本则由这些外部组织及零工群体承担。同时，平台通过平台规则、业务考核、大数据和算法等手段对外部组织及零工群体进行严密管控，以确保整个平台业务的高质量运营并降低“委托—代理”关系下的各类风险。我们将上述内容用表 1 来进行总结。

表 1　　合约关系下典型平台企业的灵活积累

平台类型	合约关系构成的管控体系	合约关系类型	平台企业的灵活积累策略
内容平台	平台+MCN 机构/公会+零工群体	1. 平台与外部专业化组织之间是外包/业务合作关系 2. 外部专业化组织与零工群体之间以劳务关系为主	1. 平台不仅基于合约/平台规则对外部组织的运营进行管控和塑造，还通过绩效考核、末位淘汰等方式督促外部组织按照平台要求组织劳动者进行灵活专业化的生产 2. 平台通过大数据和算法对劳动者的劳动过程实施监测和管理，通过动态定价实现劳动力的动态供给
电子商务平台	平台+工厂/作坊+零工群体		
物流快递平台	平台+加盟商+零工群体		
网约车平台	平台+租赁公司+零工群体		
外卖平台	平台+加盟商/外包公司+零工群体		

四、零工群体基于剩余控制权的自我保护

平台作为一个多边市场，连接着服务/商品的供给方与需求方；在各类平台中，零工群体是各类服务性劳动的供给方。尽管平台通过与零工群体缔结合约关系，对零工群体的劳动过程进行了详细的拆分及标准化要求，并对零工群体的劳动结果有着精细化的规定；但这些合约内容无法穷尽零工群体与客户之间的微观互动，也无法穷尽零工群体所面临的复杂多变的劳动空间与劳动场景。因

此，零工群体享有不同程度的剩余控制权，且会充分发挥主体性进行自我保护。在与平台的互动中，零工群体会积极争取自身利益，并积极应对劳动过程中的不确定性；在与客户的互动中，零工群体主要采取陌生关系熟悉化的策略来为自己争取利益和灵活的劳动空间；在零工群体内部，零工也会积极团结起来并进行自我主体性的建构。

（一）零工群体与平台、客户的互动

平台的发展与数字技术的完善是一个渐进的过程，平台规则、大数据和算法无法完全覆盖劳动者复杂多变的劳动场景，平台与零工群体之间存在利益博弈。以某网约车平台为例，在平台算法的剥削下，基于地缘关系的民间车队会联合采取一系列抗争策略以争取自身利益（丁未，2021）。以某外卖平台为例，尽管平台的抽象算法与治理技术试图将骑手对象化为规定动作的完成者，接单、取餐、送餐这些规定步骤也在劳动中被明确拆解出来；但是，伴随着劳动空间的扩展以及更多不确定因素进入劳动秩序之中，骑手与管理方之间呈现出越来越多的协作关系，“共谋”应对社会空间的种种挑战（付堉琪，2021）。

平台重塑了传统服务劳动的交易匹配方式。例如，传统的家政服务、中介经纪服务更多是嵌入强人际关系网络中的，交易中体现的是人际关系的发展和运用，交易匹配也是通过多次博弈和协商达成的，经济行为嵌入社会关系的特征明显；但是对于平台上的零工群体，其服务对象在空间上更加广泛和分散，交易匹配也是在“以消费者为中心”的平台规则主导下完成的，因此零工和服务对象的关系是陌生的、精于计算的。在服务关系逐渐陌生化、理性化的背景下，平台上的零工群体会采取“陌生关系熟悉化”的行动策略以争取平台控制之外的自由行动余地。

以物流快递平台为例，面对平台企业严格的管理制度和客户多样化的服务需求，快递员在情感劳动过程中逐渐摸索出多种应对策略。他们能够快速区分客户类型、建立熟人关系并提供差异化的情感劳动，这不仅降低了客户投诉的风险，也缓解了快递员的身心压力（张杨波，2021）。尽管有制度监管及技术创新来控制快递员的具体劳动过程，但是快递员在配送的路上发展出常规路径、例外路径和循环路径三种配送路径，在配送方式、客户关系方面也与不同类型的

客户形成了正式合同、逆向合同、虚化合同、代管合同和托管合同五种模式（张杨波，2020）。快递员通过“熟络”、固定化派送区域、互惠等策略，在社区中将市场化的关系构建成一种拟朋友关系，把陌生人纳入这种关系网络中，从而使自己成为城市社区中“熟悉的陌生人”，最终提高自己在劳动过程中的自主程度（庄家炽，2022）。

再以家政服务平台为例，在中国熟人社会中，拟亲属化关系的建构特征以及基于礼物交换的人情建构使关系兼备情感性和工具性。在这一文化背景下，家政工得以在劳动过程中通过与客户的互动建立起具有真挚性、延续性和独立性的自然情感模式。这是一种互惠互利的工作模式（梁萌、李坤希、冯雪，2022）。在直播平台中，“关系劳动”的现象更加明显。主播根据自己与观众在直播间内外形成的经济关系与亲密关系，践行“标价”“区分”和“界限”三种行动策略（董晨宇、叶蓁，2021）。

（二）零工群体内部：“团结起来”与主体性建构

尽管数字技术成为平台劳动控制的手段，但数字技术也同时成为零工群体社会团结和自我主体性建构的新工具。网约车平台的司机会基于微信群实现社会联结、形成主体表达并获取社会支持（庄曦、王旭、刘百玉，2019）。货运平台的司机借助新媒体技术的赋权，跨越数字鸿沟，通过微信群联结，使得原子化的个体实现了再组织化；然后通过灵巧的移动协调动员机制，使得集体利益转化为集体行动，建构出情感认同的虚拟团结，进一步衍生出救援、议价等实用团结，最终一种新型团结经济得以建构（刘楠，2021）。从虚拟团结到实用团结的关键，在于线下关系与线上关系的联结。货车司机通过对线上线下社会关系的复刻式整合以及社会性激励机制的建设，使虚拟社群得以克服广泛性易得、行动力难寻、自主性有余、合法性不足等弊端（刘学、任奕飞、刘金龙，2021）。

再以媒介内容类平台为例，在网络文学写手这一类数字零工群体中，平台会对写手采取产量竞赛、创意规训、合约上的权责置换等多重管控。面对平台的严苛管控，写手会调试意识形态来适应这种处境，也会采取个体抗争或社群联结的策略来降低职业发展的不确定性，赢得更多的职业发展机会（张铮、吴福仲，

2020)。在某视频内容平台中,UP 主通过内容浏览、视频上传与交流讨论,完成了人际关系的建立、技能和知识的获得。在亚文化社区氛围的感召下,UP 主通过文化生产来建立自己的话语体系,在与主流文化的对抗中建构自我主体性与社会认同(孙冬鑫、刘鸣筝,2021)。

五、结论与讨论

当前,平台企业已经成为各个学科的热点研究对象,但是平台企业后福特主义式的灵活积累机制尚未得到充分重视,也缺少理论解释。本文从新制度经济学合约理论的视角解释了平台企业的灵活积累机制。这一灵活积累机制的运作基础是平台企业与大量外部专业化组织及零工群体建立的不平等合约关系。这一灵活积累机制的运作过程分为两个方面:其一,在不平等的合约关系下,平台得以转移规模化扩张所带来的大部分固定资产成本、人力成本和管理成本至外部专业化组织及零工群体,从而实现轻资产化运营;其二,平台为规避合约关系中的"委托—代理"风险,会综合运用平台规则、业务考核、大数据和算法等手段,对外部专业化组织及零工群体构筑严密的管控体系,并严格规定产品/服务的生产过程和结果。当然,由于合约关系的不完全性以及合约关系中剩余控制权的存在,零工群体也会积极采取各种行动策略维护自身利益、构建自身主体性。这一理论解释适用于各类平台(如内容平台、电商平台、各类服务性劳动平台)。

在这一理论解释的基础上,本文也揭示了平台企业市场竞争优势的关键基础之一:平台企业能够在轻资产、低成本投入、低风险承担的条件下快速实现规模化扩张和资本积累,能够随市场环境变化而快速调整商业模式,这些竞争优势是平台企业独特的生产组织方式——灵活积累——机制所带来的。

当然,本文的分析主要建立在现有研究和日常观察的基础上,对平台企业灵活积累机制的探讨还缺少基于具体案例的"过程—机制"分析,这是未来需要努力的方向。此外,一些问题有待进一步研究和讨论:平台企业践行灵活积累机制的发生过程是怎样的?这一机制在不同的数字平台上又有怎样的差异?导致差异的原因是什么?在这一机制的形成过程中,平台与各类外部专

业化组织及零工群体间的互动博弈过程是怎样的？合约关系是如何达成和变迁的？等等。

参考文献

常珂璨，2018，《网络主播的职业身份认同研究》，广州大学硕士学位论文。

陈龙，2020a，《“数字控制”下的劳动秩序——外卖骑手的劳动控制研究》，《社会学研究》第 6 期。

陈龙，2020b，《游戏、权力分配与技术：平台企业管理策略研究——以某外卖平台的骑手管理为例》，《中国人力资源开发》第 4 期。

陈龙、韩玥，2020，《责任自治与数字泰勒主义：外卖平台资本的双重管理策略研究》，《清华社会学评论》第 2 期。

陈威如、余卓轩，2013，《平台战略：正在席卷全球的商业模式革命》，北京：中信出版社。

邓智平，2021，《“接单游戏”与平台经济中的劳动实践——以外卖骑手为例》，《求索》第 3 期。

丁未，2021，《遭遇“平台”：另类数字劳动与新权力装置》，《新闻与传播研究》第 10 期。

董晨宇、叶蓁，2021，《做主播：一项关系劳动的数码民族志》，《国际新闻界》第 12 期。

范璐璐、黄岩，2021，《新型柔性专业化——以平台为中心的服装业生产组织与劳动关系》，《学术研究》第 5 期。

付堉琪，2021，《“系统困不住”的外卖骑手——劳动场景视角下的变通与协作关系研究》，《新视野》第 6 期。

傅绍文、邓秋云，2004，《剩余控制权理论综述》，《经济学动态》第 11 期。

傅瑜，2013，《中国互联网平台企业竞争策略与市场结构研究》，暨南大学博士学位论文。

黄贺铂，2021，《他者的透视：MCN 机构与直播平台的想象可供性及反思》，《国际新闻界》第 12 期。

蒋淑媛、黄彬，2020，《当“文艺青年”成为“数字劳工”：对网络作家异化劳动的反思》，《中国青年研究》第 12 期。

蒋文妍，2020，《网红成名的想象——网红养成机制的现有模式及困境调研》，南京大学硕士学位论文。

赖祐萱，2020，《外卖骑手，困在系统里》，https://mp.weixin.qq.com/s/Mes1RqIOdp48CMw4pXTwXw。

李强、徐兴荣、刘凤杰，2022，《网约配送平台外包式用工关系及其形成：基于平台企业策略选择的视角》，《社会发展研究》第 4 期。

李胜蓝、江立华,2020,《新型劳动时间控制与虚假自由——外卖骑手的劳动过程研究》,《社会学研究》第 6 期。

梁萌,2017,《强控制与弱契约:互联网技术影响下的家政业用工模式研究》,《妇女研究论丛》第 5 期。

梁萌,2021,《买卖不成仁义在:互联网技术影响下平台家政小时工的社会网络变迁》,《新闻与传播研究》第 S1 期。

梁萌、李坤希、冯雪,2022,《资本之矛与劳工之盾——我国家政工情感劳动的本土化模式研究》,《社会学研究》第 2 期。

刘蒙之、刘战伟,2021,《边缘与游离:非虚构写作者的职业身份认同研究》,《媒介批评》第 1 期。

刘楠,2021,《从“虚拟团结”到“实用团结”:基于卡协青年司机微信群组织的研究》,《中国青年研究》第 1 期。

刘倩,2023,《“平台—中间商—劳动者”视角下直播公会劳动实践研究》,《中国青年研究》第 11 期。

刘善仕、裴嘉良、葛淳棉等,2022,《在线劳动平台算法管理:理论探索与研究展望》,《管理世界》第 2 期。

刘亭亭、杨晓兵,2022,《平台“再中介者”与主播“孵化器”:网络直播公会的日常实践研究》,《中国青年研究》第 4 期。

刘晓丽、王志勇,2020,《C2M 电商平台价值共创过程及升级演化机制分析》,《财会月刊》第 22 期。

刘学、任奕飞、刘金龙,2021,《虚拟社群何以组织化? ——以卡友驿站平台为例》,《新闻与传播研究》第 S1 期。

刘战伟、李媛媛、刘蒙之,2021,《平台化、数字灵工与短视频创意劳动者:一项劳动控制研究》,《新闻与传播研究》第 7 期。

刘战伟、刘蒙之、李媛媛,2022,《从“赶稿游戏”到“老板游戏”:互联网平台中自由撰稿人的劳动控制》,《新闻与传播研究》第 1 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。

牛天,2022,《数字灵工的内卷化困境及其逻辑》,《中国青年研究》第 3 期。

帕克、埃尔斯泰恩、邱达利,2017,《平台革命:改变世界的商业模式》,志鹏译,北京:机械工业出版社。

任桐、姚建华,2022,《平台经济中的“数据劳动”:现状、困境与行动策略——一项基于电竞主

播的探索性研究》,《国际新闻界》第 1 期。

邵占鹏,2022,《网络零售平台经济中的生产资料所有制》,《开放时代》第 2 期。

沈晓峰、袁钉,2020,《快递行业前瞻:加盟制如何破局》,https://baijiahao.baidu.com/s?id=1686314915688168870&wfr=spider&for=pc。

孙冬鑫、刘鸣筝,2021,《大众文化视角下数字劳动的主体性解读——基于哔哩哔哩弹幕网 UP 主的劳动过程分析》,《劳动哲学研究》第 2 期。

孙萍,2019,《“算法逻辑”下的数字劳动:一项对平台经济下外卖送餐员的研究》,《思想战线》第 6 期。

孙兆阳,2013,《劳动控制与抵抗:西方劳动过程理论评述与启示》,《中国人力资源开发》第 15 期。

涂永前、熊赟,2019,《情感制造:泛娱乐直播中女主播的劳动过程研究》,《青年研究》第 4 期。

汪旭晖、张其林,2016,《平台型电商企业的温室管理模式研究——基于阿里巴巴集团旗下平台型网络市场的案例》,《中国工业经济》第 11 期。

汪旭晖、张其林,2017,《平台型网络市场中的“柠檬问题”形成机理与治理机制——基于阿里巴巴的案例研究》,《中国软科学》第 10 期。

王茜,2020,《平台三角用工的劳动关系认定及责任承担》,《法学》第 12 期。

王蔚,2021,《数字资本主义劳动过程及其情绪剥削》,《经济学家》第 2 期。

王勇、邓涵中,2017,《论企业的交易属性》,《经济学家》第 2 期。

王勇、戎珂,2018,《平台治理:在线市场的设计、运营与监管》,北京:中信出版社。

吴清军、李贞,2018,《分享经济下的劳动控制与工作自主性——关于网约车司机工作的混合研究》,《社会学研究》第 4 期。

谢富胜、吴越,2019,《零工经济是一种劳资双赢的新型用工关系吗》,《经济学家》第 6 期。

徐林枫、张恒宇,2019,《“人气游戏”:网络直播行业的薪资制度与劳动控制》,《社会》第 4 期。

杨善奇、刘岩,2021,《智能算法控制下的劳动过程研究》,《经济学家》第 12 期。

叶韦明、金一丹,2021,《平台·公会·主播:不确定数字产业中的生产组织》,《国际新闻界》第 12 期。

袁庆明,2021,《新制度经济学教程》第 3 版,北京:经济管理出版社。

翟秀凤,2021,《多频道网络(MCN)机构背后的平台循环创新机制》,《新闻记者》第 8 期。

张雪娇、肖潇,2021,《自由灵活与隐蔽剥削:平台经济劳动关系属性的再思考》,《改革与战略》第 11 期。

张杨波,2020,《服务合同约束软化:一个快递配送模式经验观察——以安湘市速安快递公司

桥东点部为例》,《兰州大学学报》(社会科学版)第 1 期。

张杨波,2021,《重绘情感劳动图景:视角转换与路径延伸》,《新视野》第 5 期。

张铮、吴福仲,2020,《创意流水线:网络文学写手的劳动过程与主体策略》,《中国青年研究》第 12 期。

赵磊、韩玥,2021,《跨越企业边界的科层控制——网约车平台的劳动力组织与控制研究》,《社会学研究》第 5 期。

中国信息通信研究院,2021,《平台经济与竞争政策观察(2021 年)》,http://www.100ec.cn/index/detail-6595629.html。

庄家炽,2022,《熟悉的陌生人:快递工人劳动过程研究》,《中国青年研究》第 9 期。

庄曦、王旭、刘百玉,2019,《滴滴司机移动社区中的关系结构及支持研究》,《新闻与传播研究》第 6 期。

Arrow, K. 1984, "The Economics of Agency." in J. Pratt & R. Zeckhauser eds., *Principals and Agents: The Structure of Business*, Cambridge, MA.: Harvard University Press.

Hart, O. & B. Holmström 1987, "The Theory of Contracts." in T. Bewley & T. Bewley eds., *Advances in Econometrics: Fifth World Congress*, Vol. 1, Cambridge: Cambridge University Press.

Hart, O. 2017, "Incomplete Contracts and Control." *American Economic Review* 107(7).

Jessop, B. 2005, "Fordism and Post-Fordism: A Critical Reformulation." in A. Scott & M. Storper eds., *Pathways to Industrialization and Regional Development*, New York: Routledge.

Kuhn, K. & A. Maleki 2017, "Micro-Entrepreneurs, Dependent Contractors, and Instaserfs: Understanding Online Labor Platform Workforces." *Academy of Management Perspectives* 31(3).

Roobeek, A. 1987, "The Crisis in Fordism and the Rise of A New Technological Paradigm." *Futures* 19(2).

Schoenberger, E. 1988, "From Fordism to flexible Accumulation: Technology, Competitive Strategies, and International Location." *Environment & Planning D Society & Space* 6(3).

Sun, P., J. Chen & U. Rani 2023, "From Flexible Labour to 'Sticky Labour': A Tracking Study of Workers in the Food-Delivery Platform Economy of China." *Work, Employment and Society* 37(2).

Watkins, P. 1991, *Knowledge and Control in the Flexible Workplace: EAE610 The Changing*

Workplace, Victoria: Deakin University Press.

Wood, A., M. Graham & V. Lehdonvirta et al. 2019, "Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy." *Work, Employment and Society* 33(1).

信息逆差:代码权力膨胀与公众权力让渡

施颖婕*

摘要:在与算法阶层信息的交易中,互联网用户让渡易被解读、引导和操控的信息(包括各类身份、偏好、言论及动态信息)以换取所欲信息,这种不平等的信息交换将其置于信息逆差境地。公众从信息顺差地位走向信息逆差地位的过程,正是互联网从赋权转向集权的"技术倾向性倒置"社会化过程。信息逆差表现为认知、意识及行为三个方面的信息权力让渡:在认知维度,信息过滤与信息推荐形成的人工壁垒引致认知局限,资讯轰炸下的注意力中断与大脑机能超荷导致注意力失焦、记忆滑坡等认知障碍;在意识维度,对"奶瓶乐"喂养的成瘾导致个体的信息能动性降低;在行为维度,数字审查强化了个体的自我规训与主动服从,标准化、自动化的数字暴力带来更高的强制力。

关键词:信息交易　信息逆差　权力让渡　技术倾向性倒置

一、信息逆差:从"信息消化"到"信息换取"

在工业革命和市场经济完成人类文明从暴力权力(muscle power)向资本权力(capital power)的转变之后,信息革命和知识经济则被预言或将实现人类文明从资本权力到知识权力(knowledge power)的再次跃迁(Toffler, 1990)。相应地,作为向公众提供迅捷知识渠道与充量信息资讯的技术,互联网被期待将权力分配去中心化并赋权于公众。但反观现实,互联网似乎并未带来民众理性自由的增加,个体从未如现在这样可以轻易地被认识、引导与操控(Susskind, 2018)。

对此,我们往往聚焦于互联网信息提供者的身份去思考这一影响,利用"信息茧房""回音壁"等理论以解释为何信息越多却自由越少。然而,在强调信息

* 施颖婕,复旦大学新闻学院。

消化效应的同时,我们似乎容易忽略互联网在信息流动中的双重身份。民众在作为信息消费者接收互联网信息的同时,也通过互联网使用向商业公司和政府让渡出使自身更易被解读和影响的内部信息。由于互联网兼具信息供给方这一显性角色与信息收集方这一隐性角色,个体对互联网信息的消化障碍只是上述困境的部分成因,更符合全景的、深层的原因可能涉及个体在与算法阶层信息交易中的劣势地位。

在这场信息换取中,一方面,算法阶层利用数据垄断、数据自缚、数据欺骗等信息特权,在很大程度上限制了普通民众从互联网上获取知识的范围与深度,使个体获取的信息愈发碎片化和窄化;另一方面,个体在接受互联网服务的过程中,无意间出让了大量重要的个体信息,包括客观的身份信息与主观的消费偏好、认知倾向等。这些重要的隐私数据的暴露,极大程度上增加了个体的可解读性,从而使个体更容易被预测、引导和操控。

可以说,数字技术用看似庞大浩荡的信息供给换取了用户重要的个人信息。由于前者是显性输送的,后者是隐性输送的,我们常常被这场交易中一面的信息增量所迷惑,忽视了这场交易的结果其实是一种信息逆差。借由认知、意识、行为三个维度,这种信息逆差地位带来的权力博弈结果是公众权力的让渡。

二、认知维度:人工信息壁垒与信息处理机能

互联网将原先自上而下的金字塔型话语结构变得扁平化,个体在成为有力的信息生产者与传播者的同时,获取信息的途径得以多维化。因此,在互联网的海量信息环境中,用户似乎对知识触手可及。然而,一方面,在信息获取维度,用户所接收到的信息大都是其被允许看到的或是被鼓励看到的那部分内容;另一方面,在信息处理维度,人类的生物机能决定了即时呈现、数量庞大的网络信息给用户带来的可能并非骤增的真相与智慧,而是焦虑、紧张及有意识的短路。

(一) 信息过滤与信息推荐等感知控制下的认知局限

斯塔尼斯瓦夫·莱姆(S. Lem)曾在《技术大全》中预测,当人类社会发展至信息巅峰时,每一种文明都会为自己创造出一种人工的信息环境,形成一种有

助于堵住多余信息的文明胞囊(莱姆,2022)。反顾当下的信息环境,这种推断似乎并不夸张。互联网已经被用于更精确、更广泛地控制我们传递和接收的信息,从而控制我们感知世界的方式。

其中,信息过滤决定了用户被阻止看到什么。它内嵌于搜索显示、通信等诸多信息传播环节。当用户利用互联网搜索内容时,搜索引擎决定了某项查询将被哪些结果优先响应,哪些结果排名靠后而类似于不存在,哪些结果被丢失了以至于无从了解(Epstein, 2016)。① 即便用户在进行社交软件上的内部隐私交流时,也会受到过滤器的监视与影响。通信技术受制于基于禁用词的即时审查已经成为一种广泛趋势。

信息推荐则影响了用户容易看到什么、容易重复和灌输哪些主题的内容。叙事的重复将带来相关知识与信念的内化,这是因为人们对信息真实性所做出的判断是建构出来的,信息的曝光程度越高,其被用户接纳的可能性越高。相关研究发现,人们依赖三种推断方式判断信息的真实性:基础比率(base rates)、感受(feelings)以及与记忆中信息的一致性(memories)。其中,基础比率是指人们倾向于从可能性的角度去预设信息是否真实。由于人们认为所见的信息大多数都是真的,因此当人们看到自己并不了解或看不懂的信息时,会倾向于认为它是真的(Chater, Tenenbaum & Yuille, 2006);"感受即信息"理论(feelings-as-information theory)指出,人们会将自己的主观体验当作事实证据(Schwarz, 2001);最后,人们根据与已有知识和存储在记忆中的源信息的一致性进行推断(Unkelbach & Rom, 2017),相比于只呈现过一次的信息,人们更可能相信呈现过三次的信息为真。通过上述三种机制,重复会加强信息的真实性感受,从而呈现出"虚幻的真实"(illusory truth)(Brashier & Marsh, 2019)。

综上,信息过滤与信息推荐作为数据垄断的两个维度,同时控制着社会的信息流动,以至于社会的常识性认知、道德认知、社会感知都受其引导。通过让人们确信自己的某些想法是错误的、可耻的、孤立无援的,从认知层面阻止人们越轨;通过让人们相信一些理念与做法是光荣的、道德的、普遍的,从而鼓励其遵从相应的行为。即使人们出于信息的"留白恐惧症"(horror vacui)(Swirski,

① 90%的点击产生于搜索结果的第一页,因此,排名太靠后的信息在用户视域中等于不存在。

1997)而不断向大脑填充知识与资讯,也只是搬运了被挑选过的“事实”与“真理”。

(二)信息超荷下的认知障碍

除了上述人工信息壁垒所带来的选择性供给问题,海量的信息烟尘对人们的知识消化构成了生物机能上的挑战,并带来紧张、焦虑等情绪,产生与信息获取初衷背道而驰的负面结果。超负荷的信息获得正在让人们丧失过去那种平心静气、全神贯注的线性思考过程,取而代之的是简单杂乱、越快越好的新思维模式(卡尔,2010)。而从信息处理效能的角度而言,信息接收并非越多越快越好。

一方面,就信息荷载本身而言,人的大脑中存在一种名为“延迟神经元”的构造,其主要功能是延缓神经元之间的神经传导,有助于大脑理解复杂的观念与自我意识,为个体对信息的消化和吸收带来秩序、协调(沃尔夫,2012)。而当沉浸在数字化资源中时,人们从缓慢的、深思熟虑的过程中被剥离出来,弱化乃至丧失评判与挖掘信息背后意义的能力。

另一方面,就信息加载的节奏而言,互联网信息往往以多线方式呈现给用户。这种旨在分散人注意力的机制成为一个中断系统,使个体在从事某项脑力任务时被频频打断。而多项脑力任务之间的切换会加剧个体的认知负荷,导致思维分散、记忆力减弱,妨碍个体的正常思考,并带来紧张、焦虑的情绪(卡尔,2010)。可以说,互联网的多端化信息加载模式提高了思考的切换成本,从而增加了人们忽视或曲解信息的可能性。

三、意识维度:“奶瓶乐”喂养与自由饱和幻象

随着互联网以技术资本(Romele, 2020)的形式被科技公司所垄断,它部分演变为榨取和售卖用户注意力的获利工具。科技公司利用数据监控、存贮和建模向广告商贩卖确定性,基于大量 AB 测试结果而不断优化的互联网产品设计影响用户行为,诱导用户朝广告商期待的方向行动。在这一过程中,用户在频繁的点击中消耗注意力来供养和奖励这个体系,沦为信息能动性低下、政治自主性缺失的“算法燃料”。

(一) 注意力榨取与工业喂养下的信息能动性流失

在 Web 2.0 的流量逻辑下,媒体与用户在供需关系中的相对位置发生更迭,媒体的主营业务从传统的向用户售卖信息转变为向用户购买注意力,通过流量变现而盈利。与此同时,双方在供求关系中的相对位置发生了变化,媒介提供的信息资讯从需要付费订阅的稀缺品转变为泛滥的廉价品,因此用户的注意力就成为商业平台竞相追逐的稀缺品。

正因如此,基于争夺用户注意力这一趋利逻辑,商业媒体将精心准备的内容投喂给用户,从而最大限度地掌握用户的注意力以获利。正如美国网飞公司(Netflix)2020 年发布的纪录片所揭示的,科技公司设计产品时的核心逻辑便是心理学上的成瘾理论,他们利用情绪钓饵,成功地让个体在积极正强化下越发陷入技术商业逻辑的圈套与操纵之中:个体不知道哪一次能够刷到合适的内容,或者说不知道下一次能刷到什么,这种随机性奖励不停地刺激着用户继续使用下去,类似于赌场中老虎机的成瘾方式,不同的是前者更为持久、精确地向机器输送了个体偏好等使其更容易被操控和塑造的信息。经济学家泰勒·考恩(T. Cowen)指出,在能够轻易获得信息的情况下,我们通常喜欢简短、支离破碎而又令人愉快的内容(Cowen, 2002)。

当下网络环境中部分负面化、极端化、泛娱乐化等倾向,正是用户信息能动性降低、公共理性丧失的表现。以负面化为例,个体对负面事物的关注往往超过正面事物,这种负面倾向(negative bias)(徐贲,2019)被商业媒体在流量运营中利用,出现了对负面标签的滥用。同时,由于负面信息比正面信息能对人产生更为持久、显著的影响,在大量用户都缺乏必要的信息素养以应对源源不断的负面信息注入时,负面信息容易给用户带来负面感知与表达的增加。这种公共理性的丧失与社会的碎片化,与自由的衰落是紧密联系在一起的(林奇,2017)。

(二) 自由饱和幻象下的政治自主性缺失

算法技术利用“算法黑箱”隐藏起权力运作的过程,并将其幻化成用户自由意志的狂欢。人们自认为在线上享受到了较线下生活更为充沛的自由,其自由意识开始变得十分淡薄(袁媛、李娜、王灿发,2023)。正如徐贲所言,担忧这种

自由意识的丧失,其实就是担忧社会中个人与权力的关系(徐贲,2019)。

尽管互联网已经并将继续扩充诸如学习、娱乐等自由,然而自由的内在形式与自治(对内在自我的掌控)密切相关。卢梭和斯坎伦(T. Sacnlon)曾指出,屈从于自身欲望的引导并非自由,让人们只受欲望的支配形同于奴隶制(Rousseau, 1968; Sacnlon, 1972)。在"奶瓶乐"喂养下,人们越来越习惯于抛弃相互竞争的行动自由,开始享受奥德修斯解下绳索走向塞壬的自由、瘾君子的自由。用户在将精力投注于娱乐化、表浅化等主张即时满足的信息消费行为时,所支付的成本是政治自主性的减退。

公民的政治主体性关乎其担负公民责任与行使公民权利的欲求与能力。当公众不关切世界危机、社会癌症等长远的问题,而只沉醉于此时此刻的情绪满足时,他们就构成了一个在意识上相互隔绝、没有共同诉求和共同行动的空壳社会。在这样的社会中,孤独的个人无力争取自由,分崩离析的群体也难以保卫自由(徐贲,2019)。这是因为,相比于获取现成信息,对社会和政治信息深层的思考、分析和判断才有可能成为一种力量和解放。

四、行为维度:数字审查与数字化武力

韦伯(M. Weber)将精确、迅速、明晰的官僚制奉为最有效的政治控制体系(韦伯,2010),而数字技术无疑正成为这台精密仪器重要的内置结构。随着互联网对社会生活的系统性渗透,赛博空间与物质世界的界限逐渐模糊,数字技术对公众的影响已然从线上空间扩张到线下空间,亦从认知层面扩张到行为层面。日渐强大的代码权力在社会治理中扮演着越发重要的角色。

(一) 数字审查下的自我规训:服从性增加

数字时代的审查会通过外在的审视达成个体的自我规训。这种规训权力深入个体内心深处,触摸他们的身体,并植入每个人的行动和态度,植入他们的话语、学习过程及日常生活,它们比武力强制或武力威胁更为全面、有效(Foucault, 1995)。

具体而言,一方面,数字对象正在"去显化",数据收集过程变得越来越隐

蔽,人们在难以注意与察觉的情况下被自动监测与采集数据(Kelly & Hamm, 2014)。同时,数字审查存在私密化的发展趋势,它越来越多地发生在此前被视为私域的地方,导致无法画出安全领域的边界线。当人们了解自己正在被观察时,这件事本身就足以让我们思考放弃规则之外的行为。因此,尽管数字审查也许尚未无孔不入,但意识到数据收集与数字审查的隐蔽性又无力辨别数字审查是否在场,这种压力驱使人们出于对不良后果的担心而时刻谨慎。

另一方面,相比于人类记忆或纸质版记录等传统的记录载体,数字审查所得到的电子信息可能永远不会失忆,对个体产生的影响能够跨越一生甚至延至后代。在人类历史的大多数时期,遗忘是常态(Mayer-Schonberger, 2010),遗忘扮演着重要的社交角色,帮助我们挥别过去的失败和尴尬,重启新的人生阶段;但在数字生活世界中,"过去从未逝去"(Faulkner, 2012),就像谷歌公司曾公开承认它保存了用户输入的每条搜索记录和他们随后点击的每一个搜索结果,尽管某次不道德的搜索、某次违背意识形态的言论暂时潜伏在某个角落,但它们可能在半个世纪后还被调用、分析甚至用于审判。

不愿承担反叛代价的人变成了一个自我监管的主体,致力于不懈的自我监视活动。个体的每一条言论、每一个行为都可能被捕捉和保存,并用于在未来某个时刻突然决定其是否被剥夺某种权利、错失某种机遇,甚至遭遇某种非难,这会迫使个体向这种不确定性投降,在发言、行动前小心地评估言行是否越轨或存在被后来社会定义为越轨的风险从而谨言慎行。

(二)数字化武力下的外部约束:强制力增加

莱辛(L. Lessing)曾指出"代码即法律"(Lessing, 1984),因为当我们与数字技术互动时,我们需要服从代码的指示。鉴于代码与法律运作方式的区别,萨斯坎德(J. Susskind)提出了一种更具启发性的表述——"代码即权力",它能让我们去做只有武力、操纵等权力形式才能让我们就范的事情,它定义了我们可以做与不可以做的事情,并通过控制个体对外界事物的感知来施加影响(Susskind, 2018)。代码权力被国家作为新的治理工具时,可能带来比以往作为治理手段的传统成文法更为强大的约束力。

其中,代码治理让传统成文法的"事前威慑,事后惩罚"转变为更具强制性

的“事前阻止”。相比于成文法利用事后惩罚以引导个体遵纪守法的选择,代码权力让个体除了遵守规则之外别无选择。正如凯尔森(H. Kelsen)所言,法律实际上并不能禁止任何事物,它只是要求官员对特定行为进行相应的制裁(凯尔森,2021);代码权力不是警告个体“请勿进入”,而是直接设置了限制使其无法进入(Lessing, 2006)。

与此同时,代码执行具备标准化、自动化的特征,比传统依靠人力层层传递执行的方式更为迅捷高效。代码生产出的指令比自然语言更为明确和标准,并不留有解释上的灰色空间供舞弊。同时,代码权力的自动化执行极大缩减了选择性、抽查式执法的局限,大大提高了侥幸漏网的难度。

五、讨论

(一) 信息逆差下的权力博弈

互联网是一个权力争夺的地带。与任何权力一样,互联网权力关乎谁是信息的掌控者,以及谁的信息被人掌控(徐贲,2019)。当公众交给互联网的信息(包括身份、偏好、言论及动态)价值大于公众从互联网获得并吸收的信息(被操纵的知识与信息烟尘)价值时,公众就进入了一种信息逆差地带。随着这种不平等交换的持续,公众权力逐步让渡于代码权力。

一直以来,我们似乎都对互联网传播抱着某种预设,认为公众是互联网的服务与解放对象,以至于常常将互联网对公众的信息供给作为思考当前传播困境的起点。然而,互联网在成为普通民众信息工具的同时,也正成为权力机构与科技公司借以扩充权力的技术资本。纵使我们尚未悲观地认为互联网真实的顾客上帝已转变为权力机构与科技公司,而个体已沦为互联网生态圈的燃料,也不该只关注这场不平等的信息交换中公众获得的信息馈赠——我们交出的比我们想象的要多。其中,公众在有意识或无意识情况下出让的个人信息,对于权力机构与科技公司解读、预测和控制个体的价值毋庸置疑;相比之下,公众所被允许、被鼓励获得的信息对于提升智识、改善其现实境遇的作用可能有限。

在信息逆差之下,公众权力让渡于代码权力的路径可以归纳为下述三种。

在认知层面,信息过滤机制所铸造的人工信息壁垒限制了个体的认知发展,信息轰炸则使个人的生物机能陷入难堪而思维涣散、情绪焦虑、难以深度思考。在意识层面,科技公司为榨取用户注意力以盈利,而向公众投喂精心设计的"信息鸦片",致使用户高度依赖信息投喂,对信息的检索与反思能力降低;在"奶瓶乐"泛滥之下,公众沉迷于即时的情绪满足与消费快感,淡漠于社会问题的理性思考、公共事务的良性参与、公民权利的坚守,逐渐丧失政治自主性。在行动层面,数字审查导致个体的自我规训,数字化武力则加强了对个体行动的外部制约。

(二)互联网的社会化:从赋权到集权的技术倾向性倒置

温纳(L. Winner)曾提出"技术倾向性"(tendency of technology)这一概念,意指任何技术都存在一定的倾向性,会让某些人得利而某些人失利,有些技术是民主的,有些技术本身就是集权性质的(Winner, 1980)。然而,正如科林格里奇(D. Collingridge)所言,"一项技术的社会后果不能在技术生命的早期被预料到"(Collingridge, 1980),互联网亦是如此。互联网进入中国社会的30年,是公众从信息顺差走向信息逆差的过程,也正是自带权力解放这一设计逻辑的技术从赋权到集权的倾向性倒置过程。

在21世纪伊始的Web 1.0时期,作为一项脱嵌于现实社会结构的新型技术,互联网被赋予了创建"数字化理想国"的期待,也确实曾带来公共讨论的短暂繁荣。门户时代初期,在网络讨论较高的经济门槛与自驱力门槛下,知识精英作为公共空间创建与经营的核心群体,汇聚诸多具有高包容度的论坛,展开对政治、哲学领域的严肃讨论与观念争鸣。高质量信息传播的智识门槛进一步阻隔了低文化层次用户的网络表达,从用户结构维度维护了这一知识精英及其追随者的文化生态圈。在这一阶段,社会权力分配的演变大致遵循了互联网去中心化的自身技术逻辑,"技术初速度"主导了互联网的影响,尽管这种信息流动的解放因为参与成本的高昂而局限在知识精英阶层。

然而,随着移动互联网技术的成熟带来网络运用近乎地毯式的普及,赛博空间中脱嵌的"天空之城"终归还是回归凡尘,互联网落地于民情、商业、政治制度等客观社会结构之中。至此,互联网"技术初速度"主导的时代被终结,在商业趋利逻辑、代码治理等"结构加速度"的叠加下产生了技术倾向性的倒置。人类

社会的复杂性曾使得权力阶层难以完全理解——因为无法实现充分控制（Hayek，1945），但数字技术极大地削弱了人类活动的不可解读性这一对抗权力的力量，来自机器传感器的海量数据正成为总数据愈加占主导性的根源以用于更为系统化、精密化的偏好分析、行为预测与引导（Kelly & Hamm，2014）。代码权力作为一种崭新的权力类型，赋予算法阶层前所未有的强大支配力量。

从这一意义上讲，当互联网技术悬浮于线下社会时，其权力解放的技术倾向性可以在部分群体中得到局部施展；而一旦技术沉降入社会去普遍地连接、影响更多人，则非但不能为更多群体带来解放，反而因其广大的用户基数而成为资本与政治竞相攻占的目标，其技术倾向性在社会化过程中被倒置。

（三）信息逆差的群体性差异

随着代码权力对个体日常生活日益系统化的影响，了解算法在平台、服务和搜索引擎中的功能与影响，并能够尽可能有意识地、批判性地与之互动，便成为个体所能触及的一种重要的数字力量。因此，对算法的感知和意识在人群间的分化会带来不同群体在与互联网信息交易中的差异化地位。新近的相关研究发现，对算法的认知与否的群体间差异正在导致新的强化型数字鸿沟，对算法驱动的推荐内容的态度及算法认知水平与受教育程度等因素显著相关（Gran，Booth & Bucher，2021）。由此可见，尽管正如法国社会学家埃吕尔（J. Ellul）曾指出的，容易被愚弄和操纵的不仅仅是信息素质较差的群体，知识分子也不例外（Ellul，1973），信息逆差成为互联网使用过程中的一项普遍特征，但信息交易中的逆差程度仍存在较为显著的群体性差异，用户底层化与信息逆差可能存在更深的关联。

一方面，当法律框架与逻辑推演框架离场、争论脱离事实基础时，诉诸情绪成为网络讨论最为便捷的途径。这也是为什么低收入、低学历群体的情绪和言论相对而言更具煽动性，更容易被“红鲱鱼”“烟幕弹”等互联网上破坏公共理性的搅局手法所牵引。另一方面，信息烟尘更加符合底层群体的信息消费偏好。投入大量精力和经济在网站上的人较小概率是现实生活充实的人，“奶瓶乐”的设计初衷就是为了安抚被现实社会边缘化的 80%的人，用令人陶醉的消遣娱乐产品填满他们的生活，转移其注意力和不满情绪，从而安定下来。

相比而言,受教育程度较高的群体更容易掌握门槛较高但更为多样化的信息途径,对信息投喂的依赖性整体较低,仍保留较高的信息能动性与主体性。尽管在“沉默的螺旋”效应下,网络话语空间中的“底层抢占”使得理性、中立的纠偏往往遭到淹没、驱逐而失语,但同时我们也应该看到,越来越多的人开始以新卢德主义(neo-Luddism)审视互联网对个体权力的影响,对“数字家长主义”的反思愈发频繁地出现在意图规避审查的语言加工、free 等利用数字锁链按时屏蔽社交媒体的手机应用、增强隐私的密码学等旨在数字救赎与技术抵抗的行为中。

参考文献

卡尔,2010,《浅薄:互联网如何毒化了我们的大脑》,刘纯毅译,北京:中信出版社。

凯尔森,2021,《纯粹法学说》,雷磊译,北京:法律出版社。

莱姆,2022,《技术大全》,云将鸿蒙、云将鸿蒙二号机、毛蕊译,北京:北京日报出版社。

林奇,2017,《失控的真相:为什么你知道得很多,智慧却很少》,北京:中信出版社。

韦伯,2010,《经济与社会》,阎克文译,上海:上海人民出版社。

沃尔夫,2012,《普鲁斯特与乌贼:阅读如何改变我们的思维》,王惟芬、杨仕音译,北京:中国人民大学出版社。

徐贲,2019,《人文的互联网:数码时代的读写与知识》,北京:北京大学出版社。

袁媛、李娜、王灿发,2023,《算法推荐时代的技术规训与价值重塑》,《新闻爱好者》第 10 期。

Brashier, N. & E. Marsh 2019, “Judging Truth.” *Annual Review of Psychology* 71(1).

Chater, N., J. Tenenbaum & A. Yuille 2006, “Probabilistic Models of Cognition: Conceptual Foundations.” *Trends in Cognitive Sciences* 10(7).

Collingridge, D. 1980, *The Social Control of Technology*, London: Frances Pinter.

Cowen, T. 2002, *Creative Destruction: How Globalization Is Changing the World's Cultures*, Princeton: Princeton University Press.

Ellul, J. 1973, *Propaganda: The Formation of Men's Attitudes*, J. Lerner trans., London: Vintage.

Epstein, R. 2016, “The New Censorship.” http://www.usnews.com/opinion/articles/2016-6-22/google-is-the-worlds-biggest-censor-and-its-power-must-bu-regulated.

Faulkner, W. 2012, *Requiem for A Nun*, Maryland: Vintage Books.

Foucault, M. 1995, *Discipline and Punish: the Birth of the Prison*, New York: Vintage Books.

Gran, A.-B., P. Booth & T. Bucher 2021, "To Be or Not to Be Algorithm Aware: A Question of A New Digital Divide?" *Information, Communication & Society* 24(12).

Hayek, F. 1945, "The Use of Knowledge in Society." *American Economic Review* 35(4).

Kelly, J. & S. Hamm 2013, *Smart Machines: IBM's Watson and the Era of Cognitive Computing*, New York: Columbia Business School Publishing.

Lessing, L. 1984, "Looking Back [Man of High Fidelity: Edwin Howard Armstrong]." *IEEE Potentials* 3(3).

Mayer-Schonberger, V. 2010, "Delete: The Virtue of Forgetting in the Digital Age." *World Future Review: A Journal of Strategic Foresight* 2(1).

Romele, A. 2020, "Technological Capital: Bourdieu, Postphenomenology, and the Philosophy of Technology Beyond the Empirical Turn." *Philosophy & Technology* (1).

Rousseau, J.-J. 1968, *The Social Contract*, M. Cranston trans., London: Penguin.

Sacnlon, T. 1972, "A Theory of Freedom of Expression." *Philosophy and Public Affairs* 1(2).

Schwarz, N. 2001, "Feelings as Information: Implications for Affective Influences on Information Processing." in L. Martin & G. Clore eds., *Theories of Mood and Cognition: A User's Guidebook*, New Jersey: Lawrence Erlbaum Associates Publishers.

Susskind, J. 2018, *Future Politics: Living Together in A World Transformed by Tech*, Oxford: Oxford University Press.

Swirski, P. 1997, *A Stanislaw Lem Reader*, Evanston: Northwestern University Press.

Toffler A. 1990, "Power Shift: Knowledge, Wealth, and Violence in the 21st Century." *Powershift Knowledge Wealth & Violence at the Edge of Century* 5(4).

Unkelbach, C. & S. Rom 2017, "A Referential Theory of the Repetition-Induced Truth Effect." *Cognition* 160.

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

大语言模型能用来给主题模型中的主题编码吗？*

——兼论大语言模型在社会科学领域应用的前景

李　代　张博伦　周浥莽**

摘要：ChatGPT 是由 OpenAI 公司开发的基于大语言模型的通用应用程序，可以用来完成自然语言处理方面的任务。社会科学的研究实践涉及大量自然语言处理的任务，因而在这方面 ChatGPT 可能有广阔的应用前景。本文以主题模型为例，探讨 ChatGPT 能否为主题模型的结果生成可信的标签。我们抽取了发表在中外社会学期刊上使用了主题模型的论文，并打乱其主题与标签的顺序，在中外网络平台进行问卷调查，由一般用户评价二者谁更可能反映了原文的主题。结果表明，一般用户对原论文给出的标签评价并没有显著高于 ChatGPT 给出的标签评价，甚至在多数主题上 ChatGPT 给出的标签获得的评价更高。这说明 ChatGPT 可以用于给主题模型的结果生成标签，给研究者加以评判。不过，这也意味着即使缺乏领域专门知识，ChatGPT 的表现在有些条件下也可能达到专业研究者的水平，为社会科学研究带来了新的挑战。

关键词：自然语言　社会学方法　标签　人工智能

2022 年，OpenAI 公司发布人工智能产品 ChatGPT，其语言交互能力几乎可以达到以假乱真的程度，引发了社会的广泛关注。ChatGPT 可能是第一个被大量普通用户认为具有“通用人工智能”（artificial general intelligence）潜力的产品。一经推出，用户就开始用它搜索信息、翻译语言、编辑稿件、撰写代码，并发现很多时候其输出结果稍加修改就可以达到令人满意的水平。ChatGPT 实际上

* 本文系中国政法大学科研创新项目（项目批准号：10822521）的阶段性研究成果，并受中央高校基本科研业务费专项资金资助。感谢范新光、周涵、徐丽娟、瞿璧瑜、李月旻等师友在本文写作过程中的帮助、建议。

** 李代，中国政法大学社会学系。张博伦，浙江大学社会学系。周浥莽，中国人民大学社会学院。

是一个易用的人机交互接口,用于与该公司训练的“生成性预训练变形器”(generative pre-trained transformer, GPT)模型进行交互。由于这类模型包含的参数达到数十亿规模甚至更大,因此被称为“大语言模型”(large language model, LLM)。对于内容生产工作而言,类似 GPT-3.5、GPT-4 这样的大语言模型能否带来革命性的影响?或许它将大大提高内容生产的效率,而同时带来失业的危机;或许它最终被证明无法克服目前的弊端(如它有时会自信地给出事实性错误的答案),因而不堪大用。这些问题有待社会科学经验研究的探索。

在展开对大语言模型的经验研究与理论反思时,社会科学可以首先考虑其对自身知识生产的影响,因为知识生产亦是一种内容生产。以 ChatGPT 为代表的大语言模型可能会给现有的社会科学研究模式带来莫大机遇,也可能带来巨大挑战。例如,通过 ChatGPT 进行对某个领域的文献综述,可以帮助初学者快速掌握关键文献;通过 ChatGPT 对论文进行文字编辑,可以使其更加文从字顺乃至符合学术话语规范。《自然》子刊《生物医学工程》的编辑表示,他们已经在用 ChatGPT 辅助编辑工作,而这也意味着他们将更加严肃地思考综述性文章的价值,对论文的语言运用也有更高的标准(Prepare for Truly Useful Large Language Models, 2023)。毕竟,ChatGPT 可以自动完成信息整合类的综述,也能用规范的语言写作。想必在不远的未来,期刊编辑将面临一个难题:如何判断自己接到的投稿是人写的还是模型生成的?既然大语言模型几乎一定会对社会科学研究产生影响,那么我们应该如何将大语言模型融入社会科学研究呢?

已有文章较为宏观地探讨了计算方法在社会科学中的应用(陈云松、吴晓刚、胡安宁等,2020;周涛、高馨、罗家德,2022),本文则试图在其基础上以小见大,尝试将大语言模型用于一个特定的社会科学研究方法上:给主题模型的结果打标签。为此,本文选取了 2013 年以来在各大中文核心期刊和主要英文期刊上发表的采用了主题模型方法的社会学论文,将 ChatGPT 所打的标签与原文中的标签进行对比,并通过在线调查探讨其说服力:一方面,这能切实地为主题模型在应用中存在的实际问题提供解决方案;另一方面,这番论述以小见大,揭示了当前社会科学研究者面临的方法上的“失控”危机。下面,本文将首先介绍主题模型方法,再陈述大语言模型用于给主题模型的结果打标签是否可信的研究过程,最后对“失控”进行分析。

一、文献综述

(一) 主题模型

主题模型(topic model)(Blei, 2003)最早由计算机科学家开发,后来被引入社会学研究。它基于词汇在语料中共同出现的频率及不同模式,将词语分为互不排斥的若干组。举例而言,体育新闻报道中“奥运会”“金牌”“训练”等词汇经常会共同出现在同一文本中,因此很可能被分入同一组。由于经常在某些文本中共同出现而分入一组,因此同一组词语往往被理解为一个具有实际意义的“主题”。主题模型尽管并不理解每一个主题是什么,但是能识别出词汇共同出现的不同模式,从而将各个潜在的主题区分、识别出来。

常用的主题模型采用潜在狄利克雷分配(latent dirichlet allocation, LDA)模型。这种模型将语料分为三个层次——文档、主题、词语——其中文档和词语是已经被观察到的。这种模型假定:主题在文档中的分布、词语在主题中的分布是两个分别由潜在狄利克雷分配随机生成的多项分布。一个文档按照一定概率产生出若干主题后,每个主题又按照一定概率产生若干词语,最终形成这一文档中被观察到的一组词语及其频率。研究者一般使用“吉布斯采样”(Gibbs sampling)来进行参数估计(Blei,2012)。

上述模型是一种“非监督”机器学习方法。“非监督”指不需要研究者向其提供已经打好标签的训练数据,模型就可以直接对文本展开分析。但是在使用主题模型时,研究者需要首先设定一系列超参数,如假设语料中存在 k 个主题,然后才能由主题模型进行拟合。k 的大小本身没有特别客观的依据,取决于研究者对文本的认识和对初步结果的评估。k 如果太小,可能会导致一个主题内部的特征词看上去不像是来自同一主题;k 如果太大,则可能会出现很多难以解读的主题,而且也会加大解读成本。

在主题模型的具体应用中,研究者如何确定一个主题是什么?令全部语料中存在的词的总数量为 j,那么一个主题就是这 j 个词以及标识这 j 个词在该主题中出现概率的向量。k 个主题并列,就得到了主题模型的输出结果之一——

一个有 k 列、j 行的矩阵，其中每一列是一个主题，每一个主题下有 j 个元素，分别表示语料中所有词在该主题中的出现概率。研究者将所有词按照在某一主题中的出现概率从大到小排列，可以通过阅读该主题中出现频率最高的词对这一主题的含义有所推测和解读。由于主题模型允许某些词在多个主题中都成为高频词，因此它比简单的词频分析更符合人类语言的特征。

在主题模型中，一篇文章并不只有一个主题。恰恰相反，主题模型的一大优势就是它允许一篇文章有多个主题，更能适应文本的复杂性。若语料中有 n 个文本，则主题模型将输出一个有 k 列、n 行的矩阵，每一行是同一个文本里包含的 k 个主题各自的比例，总和为 1。不难理解，对于这些文本而言，主题模型得到的各个主题的比例大小就是其文本的一些特征，可以用来进行进一步的分析。这相当于一种对文本特征进行降维的算法，如以词频计，原来文本有 j 个特征（因为语料包含 j 个词，一篇文章中每个词都对应着一个词频），而现在降维成了 k 个特征。

社会科学研究者在使用主题模型时，往往根据一个主题里概率或者其他统计量最高的前 10—100 个词来判断这个主题是什么。这些词被称作“特征词”，是给主题编码的基础。主题编码并非主题模型的内在部分，对用模型进行预测这一工具开发的初衷而言也并非必要（Blei，2003），但对社会科学研究却可能至关重要。这是由社会科学研究与产业界任务性质的差别决定的。例如，一家互联网企业可能会依据文档中不同主题的比例建立预测模型，用来识别不同类型的文本：甲主题比例较高的是体育新闻，乙主题比例较高的是娱乐新闻，等等。只要模型预测表现较好，甲主题、乙主题到底是什么就无关紧要，可以留在“黑箱”里。但是对于社会科学而言，研究者往往需要人为地给主题打上相对可信的标签，否则无法进一步对研究关心的理论问题展开讨论。这时，主题模型的作用在于阐释文本，并没有被用作预测模型。

例如，迪马鸠（P. DiMaggio）等人深入研究了新闻如何报道政府对艺术的支持情况。他们根据各个主题中出现频率最高的 100 个词，将 12 个主题分别归纳为“冲突”（conflict）、“地方政府项目和资助”（local government projects and funding）以及“特定的艺术流派、事件或赞助目标”（specific arts genres，events，or grant purposes）等标签。这些标签不仅具体明确，而且具有深刻的理论内涵

(DiMaggio, Nag & Blei, 2013)。

(二)主题模型的编码问题

主题模型的编码问题是不少方法论研究的焦点。在主题模型最初被引入社会学时,研究者认为它的一大重要优势在于可以利用机器自动给文本进行编码,从而尽可能地规避编码者对编码过程的人工干预(Mohr & Bogdanov, 2013)。在主题模型出现之前,研究者如果要对大量文本进行编码,首先需要通读文本来建立对文本统贯的理解,进而将这一理解整理成指导后续编码的规则手册。也就是说,原来的编码流程是"诠释,然后计数"(Mohr & Bogdanov, 2013)。主题模型则扭转了这个顺序,它先会输出一个结果给研究者,研究者再赋予这些结果以相应的标签。这么一来,主题模型"使得研究者可以在将他们的前见用于分析之前就发现文本结构"(Mohr & Bogdanov, 2013)。这个立场也被很多数码人文学者所接受(Heuser & Le-Khac, 2012; Meeks & Weingart, 2012)。

但是在这个过程中,各个主题的编码由人工给出,存在大量的自由裁量空间。首先,研究者在给这些主题打标签时,可能带入"确认偏误"等主见。其次,有时主题模型输出的结果一时难以解读,可能是无意义的垃圾主题(Marshall, 2013),也可能这些陌生的结果其实是具有理论意义的新异主题,这恰恰是主题模型的迷人之处(Buurma, 2015)。总之,有必要对研究者凭借直觉给主题打的标签进行评估,如将主题标签同外部事件进行对比,通过文本细读进行专家判读,或者通过众包对主题的关键词与标签之间的关系进行质量判别(Grimmer & Stewart, 2013; Ying, Montgomery & Stewart, 2021)。这些策略的问题在于其成本相当高昂,应用规模难以扩大。在这个问题上,大语言模型提供了一个成本相当低而效果可能足够好的解决方案。

(三)大语言模型与主题模型

要理解大语言模型,首先要从更基本的语言模型讲起。一个相当简化的语言模型是"词袋"(bag-of-words),它认为文本的含义可以由组成它的词和词组的频率来表示。每一个词都是"独热"(one-hot)向量,是与其他词完全不同的特征。所谓"独热",指的是如果一个数据集有三个维度,那么[1,0,0]、[0,1,0]或

[0,0,1]就是独热向量,每两个独热向量之间的余弦相似度为 0。在这一模型中,词的顺序没有意义。例如,“喜欢”和“喜爱”没有关系,“我—喜欢—猫”和“猫—喜欢—我”没有区别。主题模型背后的语言观正是“词袋”。

在此基础上,研究者希望能将词之间的关联也加入考量,如词嵌入模型(word embedding)。词嵌入模型的思想是将一个词语的各种意涵投射在一个超高维度空间,再降维成一个人类可解读的多维向量。结果是,意涵相似的词语会在多维空间中处于相近位置,词语之间的关系(如反义、等级、顺次)可以通过其在特定维度上的向量关系表现出来。词嵌入模型的两种常见算法(Word2Vec 和 GloVe)都基于词与上下文的共现关系,改变了“词袋”语言模型对词语关系与顺序的漠视。

词嵌入模型进行文本分析时,可以处理不同社会范畴与阶级、性别、种族等社会类别之间的对应关系(比如“戏曲—流行乐”与“老年—青年”之间的对应关系)。例如,对过去百年间数百万种出版物进行词嵌入模型分析,会发现文化范畴的阶级属性在过去一个世纪间保持相对稳定(Kozlowski, Taddy & Evans, 2019)。已有相当数量的学者使用了词嵌入模型进行社会学研究(Stoltz & Taylor, 2021; Li, Wu & James, 2020; Caliskan, Bryson & Narayanan, 2017)。

然而,词嵌入模型本身在处理较“小”的语料时却存在较大的偏差。因此,词嵌入模型必须依赖预训练的大语言模型才能达到较好的效果。这些模型的代表是生成式预训练模型、ELMo(embeddings from language models)模型、BERT(bidirectional encoder representations from transformers)模型。这三个模型不仅将上下文的关系加入训练中,而且还加入了特定的记忆力/注意力机制。

大语言模型的规模相当惊人。GPT-1.0 是基于 BooksCorpus 数据库进行训练的,拥有 1.17 亿的参数;BERT 在这个数据库的基础上增加了英文维基百科的数据,参数量增加到了 3.4 亿。研究者发现,在预训练范式中,模型的规模、训练语料的规模与模型最终在各个任务上的表现都有着非常直接的关联(Delvin, Chang & Lee et al., 2019; Kaplan, McCandlish & Henighan et al., 2020)。Transformer 构架恰恰提供了让模型迅速变大的技术框架。于是,在 2018 年后的几年内,模型使用更多的数据,进行更多的堆叠,不断变大(Bender, Gebru & McMillan-Major et al., 2021)。ChatGPT 所基于的 GPT-3.5 Turbo 已

经拥有了2000亿的参数,并且在诸多任务上展现出了泛化的能力。比起传统的词嵌入算法,大语言模型极大地扩张了训练语料和参数的规模,使信息中更多的维度被保留下来。以ChatGPT为代表的"聊天机器人"这一交互形式,涌现出了解决复杂问题的能力。

不过,目前大语言模型还没有完全取代主题模型这类小模型。作为社会科学研究者,我们仍然希望对具有特定时空条件限制的文本进行考察,以理解文本在特定语境中的意义。那么,我们能否将大语言模型引入主题模型的应用中,帮助主题模型克服编码的困境呢?具体而言,当我们把主题模型的结果输入大语言模型之中时,大语言模型能否像研究者一样,根据它自己预训练掌握的语言模式,给主题模型的结果打上可信的标签呢?下面,我们将对这一技术的潜能进行更具体的分析。

二、数据与方法

本文的目的是评估大语言模型基于主题模型的结果进行编码的能力。如前所述,对于研究者而言,用于确定每个主题应如何编码的信息主要是主题模型输出的一个k列、j行的矩阵,研究者通过阅读每个主题中的若干特征词为其打上标签。采用主题模型进行研究的较为底线的惯例是向读者提供每个主题中靠前的特征词,以便读者判断研究者的编码是否可以接受(Ying, Montgomery & Stewart, 2021)。遗憾的是,并不是每一篇得到发表的社会学论文都能达到这一底线要求。

如果把编码的任务交由大语言模型完成,是否能得到可信的结果呢?把前面所说的每个主题下排序靠前的特征词输入大语言模型,大语言模型能否像人类解读者一样给出看似合理的标签?

为此,我们检索社会学中英文期刊发表的使用主题模型的论文,从中提取两方面信息:论文中主题模型的结果包含的特征词、原研究者给这个主题的标签(原文标签)。我们将特征词输入GPT-3.5,将生成的新标签(ChatGPT标签)与原文标签进行比对。为了评判新标签是否可信,我们通过两个在线调查平台(Credamo和Mechanical Turk)发布了匿名评估问卷,请网络用户在不知情的情

况下评价原文标签和 ChatGPT 标签哪个更有可能反映了原文的主题。

（一）原文主题与原文标签

我们检索了截至 2023 年 3 月社会学领域使用主题模型的已发表的中英文论文。2013 年 12 月，《诗学》（*Poetics*）出版了一期特刊，发表了若干篇采用主题模型进行研究的社会科学论文。这可能是社会科学领域最早集中发表采用主题模型进行社会科学研究的一期特刊。以此为起点，我们以“主题模型”（topic modeling）为关键词，检索了迄今在以下期刊发表的使用了主题模型的经验研究。中文社会学核心期刊：《社会学研究》《社会》《社会学评论》《社会发展研究》《中国青年研究》《青年研究》《妇女研究论丛》《人口学刊》《人口研究》《人口与发展》《人口与经济》《中国人口科学》。英文社会学期刊：《美国社会学评论》（*American Sociological Review*）、《美国社会学刊》（*American Journal of Sociology*）、《社会学方法和研究》（*Sociological Methods & Research*）、《社会学方法论》（*Sociological Methodology*）、《社会力量》（*Social Forces*）、《理论与社会》（*Theory and Society*）、《诗学》（*Poetics*）。

这番检索共发现 8 篇中文文献、50 篇英文文献。当然，这可能会有遗漏，如初次检索遗漏了 Nelson（2021），我们也在力所能及的范围内予以补齐。并不是每一篇论文都在正文中刊登了主题模型的结果。为了便于分析，本文仅选取在正文中刊登了输出结果并对主题进行了编码的文章进行研究。以此为标准进行筛选，得到 2 篇中文文献、18 篇英文文献。其中，有一篇英文文献处理的文本是中国历史典籍，因而语料是繁体中文（Miller，2013）；另有一篇中文文献处理的是英文文本，但用中文给主题打标签（郭台辉、周浥莽，2020）。最后，我们一共获得了 378 个主题以及相应主题的特征词和作者给予的原文标签。

（二）大语言模型与 ChatGPT 标签

在大语言模型方面，为了模拟人类阅读主题模型结果给出主题标签的过程，我们使用的是基于 GPT-3.5 Turbo 模型的 ChatGPT 对话任务。ChatGPT 采用基于人类反馈的强化学习（reinforcement learning from human feedback，RLHF），在 GPT-3.5 Turbo 模型的基础上对参数进行进一步微调。在其开发者

OpenAI 公司公布的 InstructGPT 的训练方案中,OpenAI 通过人类反馈的方式,构建了一个标注数据集和一个对于回答的答案进行排序的数据集,并通过这些人类反馈的数据对 GPT-3 模型进行了微调,训练出 InstructGPT。通过微调,InstructGPT 能比原来的 GPT 模型更好地处理人类使用者给出的提示词。ChatGPT 的训练方式与之相似。正是这个交互模式,使得 ChatGPT 展现出很强的任务泛化能力(Ouyang, Wu & Xu et al., 2022; OpenAI, 2022)。在本文写作时,GPT-4.0 已经推出,但本文作者尚未申请到使用权限。

在操作过程中,我们首先生成标准化结构的英文提示语(prompt),再通过 Python 接入 ChatGPT 提供的端口将提示语输入模型,并存储其返回的结果。使用接口是为了让电脑自动完成问答任务从而节省时间;如果不使用接口,普通用户也可以登录 ChatGPT 的界面、输入提示语、得到 ChatGPT 标签,只不过这个过程不够自动化。

由于 ChatGPT 的底层模型是概率模型,而且提示语的变化可能引起答案变化,因此研究应当忠实地记录研究过程中使用的提示语,而不是将其视为技术上的细枝末节。我们按照以下步骤输入提示语。首先,我们输入"你是一个标注主题模型输出结果的专家"(You are an expert in labeling outputs from topic models)来初始化系统信息。其次,我们说明:"你能依照关键词来帮我给几个来自一个主题模型中的主题标注标签吗? 这些主题的关键词如下:{特征词}。"(Could you please help me with labeling the following topics based on their top terms? The top terms fo a few topics are listed below:{特征词}.[①])我们又规定了对话模型应该返回的模型格式:Your answer should only consist of the best label for each topic in the following format:"Topic N:{label}" on a separate line for each topic. 由此,我们获得了 ChatGPT 给出的主题标签。如果改变提示语,确实可能导致结果变化,不过本文的焦点并非怎样的提示语最有效。限于篇幅,本文暂且不对此多加讨论。

通过这样的方式,我们让 ChatGPT 为前述所有主题给出标签。在此选出全

① 在原本的编码中就存在 fo 的拼写错误,但并没有影响我们得到结果,这里予以忠实记录而不加修改。

部三篇使用了中文语料或中文标签的论文，并将其中涉及的主题特征词、原文标签和 ChatGPT 标签展示如下（表 1）。此处主题编号之所以不连续，是因为相应的编号在原数据中被用作他用，而非缺失了主题。

表 1　涉及中文的三篇论文中主题模型的特征词、原文标签和 ChatGPT 标签

编号	论文名	特征词	原文标签	ChatGPT 标签
2	“黑箱”中的青年	抖音、微信、B 站、淘宝、小红书、知乎、京东、今日头条、朋友圈广告、豆瓣、快手	媒体平台	社交媒体和电商平台
3	“黑箱”中的青年	大数据、信息茧房、杀熟、信息壁垒、精准推荐、曝光度、流量池、监听、深陷、深度学习、神经网络	专业知识	大数据和个性化营销
4	“黑箱”中的青年	类似、历史、浏览记录、监控、偷听、热门、屏蔽、偷窥、窃听、总是、反复	运作机制	监控和隐私担忧
5	“黑箱”中的青年	广告、个性化、精准、投放、隐私、一致化、知识孤岛、定制、投送、趋于、推广	运作影响	在线广告推送和个性化
6	“黑箱”中的青年	推荐、喜欢、内容、隐私、方便、信息、感兴趣、符合、信息茧房、时间、推送、容易、娱乐、喜好、需求、侵犯、淘宝、音乐、商品、泄露、窥探、搜索、监视、有好有坏、中立、满足、兴趣、有利有弊、单一	娱乐内容	用户偏好和隐私
7	“黑箱”中的青年	信息、信息茧房、资讯、新闻、感兴趣、算法、隐私、搜索、容易、观点、全面、符合、侵犯、消息、客观、平台、兴趣、获取、浏览、同质化、重复、认知、泄露、监视、局限、热点、社会、标题、软件、垃圾、极化	资讯内容	信息隐私和在线平台消费
8	“黑箱”中的青年	广告、不喜欢、隐私、讨厌、推荐、内容、朋友圈、算法、推送、感觉、侵犯、反感、投放、精准、很烦、太多、不感兴趣、泄露、监视、软件、平台、体验、烦人、消费、厌恶、厌烦、社交、浪费时间、用户、重复	广告内容	对广告推送和侵犯隐私的负面态度

（续表）

编号	论文名	特征词	原文标签	ChatGPT 标签
10	Rebellion, Crime and Violence in Qing China	ni 拟、zui 罪、an 案、xing 刑、bu 部、shen 審、li 例、zhao 照、zhe 斩、si 死、yi 議、jiao 絞、shang 傷、ding 定、ming 命、gou 勾、jian 監、lv 律、ou 歐、sha 殺、gai 改、li 理、bi 斃、qiu 秋、shi 氏	crime	criminal justice/ system
11	Rebellion, Crime and Violence in Qing China	huo 獲、na 拏、fei 匪、yan 嚴、fan 犯、di 地、bu 捕、ji 緝、dao 盜、chi 飭、an 案、tu 徒、xian 縣、jie 劫、shou 首、guan 官、cheng 懲、qiang 搶、tao 逃、min 民、jing 經、zi 滋、du 督、zei 賊、bian 弁	unrest	crime and law enforcement
12	Rebellion, Crime and Violence in Qing China	huo 獲、na 拏、xun 訊、gong 供、an 案、jiao 教、jie 解、yan 嚴、shen 審、jiu 究、jing 經、ni 逆、li 李、xian 現、sheng 省、xian 縣、wang 王、zhang 张、liu 劉、jia 家、mi 密、shou 首、que 確、wang 往、fang 訪	sedition	legal proceedings
13	Rebellion, Crime and Violence in Qing China	zei 賊、bing 兵、fei 匪、jiao 剿、ni 逆、cheng 城、gong 攻、jun 軍、li 力、fu 復、cuan 竄、yong 勇、guan 官、jing 經、dai 帶、ji 擊、lu 路、huo 獲、sheng 勝、ke 克、du 督、bao 保、shou 首、shu 數、sha 殺	rebellion	military operations
14	Rebellion, Crime and Violence in Qing China	bing 兵、jiao 剿、jun 軍、zei 賊、fei 匪、fang 防、cuan 竄、dai 帶、li 力、chi 飭、xian 現、lu 路、bei 北、zhou 州、ni 逆、yong 勇、nan 南、sheng 省、dong 東、cheng 城、pai 派、yan 嚴、sheng 勝、xun 迅、du 督	major rebellion	military action and tactics in ancient China
16	历史社会学能化解学科之争吗	social、history、culture、politic、society、use、term、history、study、one、universal、subject、new、can、theory	理论与方法论	文化和社会研究
17	历史社会学能化解学科之争吗	one、use、term、time、new、subject、people、see、culture、content、like、work、life、history、way	文化研究	生活方式和文化

（续表）

编号	论文名	特征词	原文标签	ChatGPT 标签
18	历史社会学能化解学科之争吗	use、population、data、term、content、census、age、subject、year、rate、city、number、difference、economic、century	量化史学	人口特征和数据分析
19	历史社会学能化解学科之争吗	state、govern、new、public、policy、social、develop、nation、institution、politic、economic、city、universal、power、local	制度演变	公共政策与治理
20	历史社会学能化解学科之争吗	land、product、market、labor、use、economy、agriculture、peasant、industry、farm、rural、economy、term、capital、village	政治经济研究	农村经济
21	历史社会学能化解学科之争吗	nation、politic、state、people、core、Jewish、new、violence、movement、Jew、term、one、right、identity、group	民族/国家	犹太民族主义运动中的政治认同和暴力
22	历史社会学能化解学科之争吗	worker、class、social、politic、labor、union、work、German、revolution、industry、state、use、term、content、movement	革命/劳工	劳工与社会运动
23	历史社会学能化解学科之争吗	law、India、court、India、legal、case、century、use、state colony、one、subject、govern、British、author	法律研究	印度法律体系
24	历史社会学能化解学科之争吗	Islam、state、Muslim、Chinese、term、China、Ottoman、core、universal、use、new、subject、religious、empire、nation	亚洲研究	穆斯林帝国和中国
25	历史社会学能化解学科之争吗	politic、party、election、vote、use、democrat、state、new、term、content、republican、support、American、subject、elector	民主研究	政治和选举
26	历史社会学能化解学科之争吗	women、family、children、men、marriage、household、work、social、female、gender、male、marry、live、age、class	性别与婚姻	性别和家庭动力学

（续表）

编号	论文名	特征词	原文标签	ChatGPT 标签
27	历史社会学能化解学科之争吗	colony、slave、African、Africa、British、European、French、term、new、trade、subject、state、core、Indian、world	帝国研究	跨大西洋奴隶贸易和殖民主义
28	历史社会学能化解学科之争吗	American、black、white、school、education、new、south、universal、racial、African、race、immigration、York、work、class	种族研究	美国的民权运动与教育
29	历史社会学能化解学科之争吗	health、medical、disease、death、body、medicine、sexuality、social、practice、century、public、cause、report、new、work	卫生/健康	公共健康与医药
30	历史社会学能化解学科之争吗	Mexico、Spanish、Latin、state、term、use、new、Mexican、America、see、Brazil、Indian、city、Spain、American	拉美研究	拉丁美洲研究

（三）可信性评估

如何评估 ChatGPT 标签的可信性呢？我们设计了两个简单的调查问卷进行在线发放，请普通人对两类标签进行评估。由于论文分中英双文，我们分别在中美进行了调查。首先报告中文调查的过程。表 1 报告了两篇中文论文和一篇涉及中文语料的英文论文，我们将其结果进行整合，通过 Credamo 平台设计成电子问卷，再通过微信向联系人分发。问卷设计如下。

我们将调查题目设为“网络碎片化信息的认知”，并将调查任务伪装成评估两个人工标签。例如，在第一篇论文（赵龙轩、林聪，2022）方面，我们询问：“我们从一些长文章中每篇提取 10—20 个高频关键词，请 2 位来自不同专业的本科生仅通过阅读这些高频关键词来推断这篇文章具体的主题。您认为哪种更可能反映了原文的主旨？如果两个差不多好，请选择‘差不多’。”再下设若干道小题，每一道小题列出主题模型的特征词，让答题人选择原文标签、ChatGPT 标签或是“差不多”。为了避免选项顺序对答题人造成干扰，我们随机调整原文标签和 ChatGPT 标签出现在选项中的顺序。这样，如果答题人以完全随机的方式或

者固定选择选项的方式答题,两类标签被选中的期望概率就都是 50%。

依照同样的结构,我们继而询问后两篇论文,所用设问方式分别为:“我们收集了关于清朝中国的一些史料,统计每篇史料中的高频关键词。另请历史系的 2 位美国留学生通过这些高频关键词来推断史料的具体主题。”“我们收集了一些历史社会学论文,统计每篇论文中的高频关键词。另请社会学系的 2 位同学通过这些高频关键词来推断论文的具体主题。”

之所以不告知答题人我们旨在研究论文的原文标签和 ChatGPT 标签,是为了避免答题人受此干扰。如果答题人对 ChatGPT 这类技术有一定的立场,则可能会因为有标签是 ChatGPT 给出的而改变评估标准。需要指出的是,“从长文章中提取 10—20 个高频关键词”“通过关键词来推断这篇文章具体的主题”,并不符合主题模型实际运作的机制。主题模型允许一篇文章有多个主题,允许主题共享高频特征词,研究者在给主题打标签时也并不仅有特征词这一信息。但是,如果完全依照主题模型运作的机制询问答题人,对于不熟悉主题模型的答题人来说理解的难度太大,而且对我们的研究目的来说恐怕也没有什么帮助。因此,我们采取了“通过高频关键词还原文章主题”的故事,希望答题人能够轻易理解其任务而又不偏离我们的研究目标。

该研究共回收 33 份问卷。由于研究者的社交网络有一定特殊性,因此答题人以在校本科生和硕博研究生、人文学科或社会科学背景者为主。由于该研究并非一般意义上的社会调查,也没有可信的理由认为受访者的社会经济地位等属性会显著影响其评估工作,因此受访者的个人情况尽管不具有对总人口的代表性,也不妨碍调查结果的可信性。

英文调查在 Mechanical Turk 众包平台进行。与 Credamo 不同,该平台可以从其用户库中调用答题人,不需要研究者自行通过社交网络寻找答题人。我们设计的调查任务结构与中文调查相似,在此不再赘述。

由于在这个平台上找人答题需要支出费用,因此我们限定了英文文章主题的数量。首先排除前面已经调查过的使用中文语料的英文文章。对于剩余的 17 篇英文论文,我们首先排除了一篇语料是意大利语的文章。对于 2021 年 9 月之前发表的 13 篇文章,我们通过等权重的简单随机抽样方式抽取了 8 篇文章,再保留 2021 年 9 月之后发表的 2 篇文章。之所以要考虑 2021 年 9 月这一

断点,是因为 GPT-3.5 的训练数据截止到 2021 年 9 月,我们想知道对于一定在其训练数据之外的论文,其表现是否有差别。对于这 10 篇文章,如果文章报告的主题数量少于或等于 10 个,我们抽取全部的主题;如果文章报告的主题数量多于 10 个,我们从中等权重随机抽取 10 个主题。由此,我们获得了 89 个主题。按照与中文调查类似的结构,我们采取伪装、随机调换标签顺序、生成调查任务的方式,并在该平台上发布了问卷,每份问卷请 30 个答题人进行回答。

由此,我们一共评估了 3 篇涉及中文的文章(27 个主题)、10 篇英文文章(89 个主题)中主题标签的情况。如果答题人选择原文标签的比例并不比选择 ChatGPT 标签的比例高出太多,则很可能说明 ChatGPT 标签足以以假乱真。本文约定,如果超过半数的答题者选择原文标签,则说明原文标签比 ChatGPT 标签更令人信服。或者,如果选择原文标签的人数比选择 ChatGPT 标签的人数更多,则说明原文标签更令人信服。反之,则说明 ChatGPT 标签足以以假乱真。

三、结果

(一) 整体结果

在中文文章的 27 个主题中,只有 1 个主题有等于或超过 50%的受访者选择了原文标签,1 个主题选择原文标签的人数超过选择 ChatGPT 标签的人数;在英文文章的 89 个主题中,只有 5 个主题有等于或超过 50%的受访者选择了原文标签,27 个主题选择原文标签的人数超过选择 ChatGPT 标签的人数(图 1)。换句话说,不论以前面约定的哪个标准来判断,在大多数主题上,ChatGPT 标签都达到了以假乱真的程度。还有一个主题比较特殊,就是 Goldenstein & Poschmann (2019)中的一个主题其两个标签一样,因而答题人的选择不能说明问题。但即使排除这个特殊的主题,剩下的数据也仍然支持前面的观察。

(二) 自变量的潜在影响

对 GPT-3.5 的训练只使用了 2021 年 9 月之前能够获得的英文文本数据,这构成了这一模型的知识阻断点。该模型在某些主题上表现较好,可能是因为

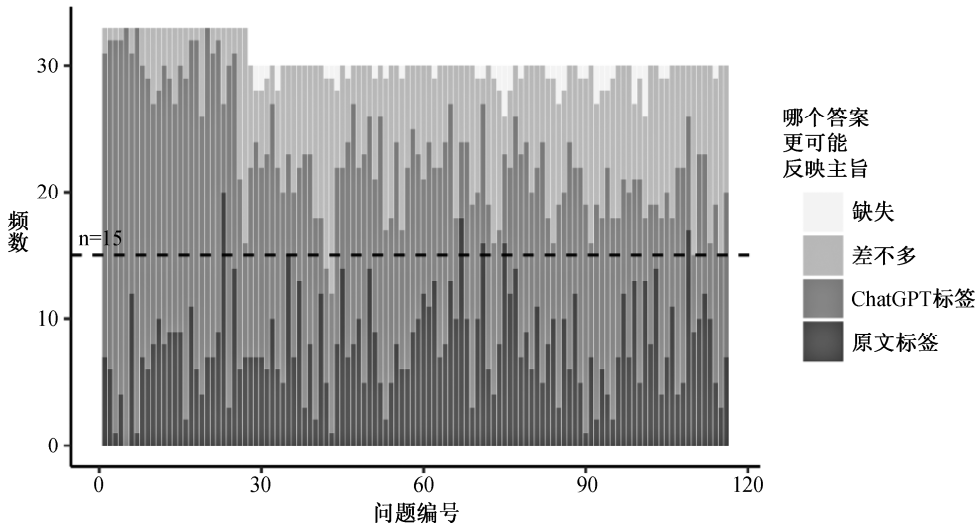


图 1　受访者在原文标签与 ChatGPT 标签之间选择情况总览

数据来源：Credamo 和 Mechanical Turk 正式调查数据。

这些发表于 2021 年 9 月之前的论文在其训练数据当中，因而打标签对模型来说不过是简单的信息检索过程。为此，我们把不可能存在于训练数据的 2 篇中文论文、2 篇英文论文标注出来，观察模型表现是否有显著差别。结果表明，是否存在于训练数据中对模型表现的影响并不显著。我们还比较了 3 篇涉及中文的文章和其他纯英文文章的表现，也未发现显著差别。由于“不可能在训练数据中”“涉及中文”这两个属性高度重叠，这里仅报告前者的结果（图 2）。

为了加强对描述统计的信心，我们以每一个主题为分析单位，以“选择原文标签的比例”为因变量，“是否可能在训练数据中”“是否涉及中文”为自变量，进行逻辑斯蒂回归（由于因变量是取值从 0—1 的变量，不符合线性回归对因变量的假设，因此应采取逻辑斯蒂回归），结果表明变量系数均不显著，印证了这里的观察，结果不赘述。

（三）对上述结果的讨论

如果 ChatGPT 标签足以以假乱真是因为其标签与原文标签高度相似也就罢了，但令人惊讶的是，很多时候 ChatGPT 标签与原文标签有明显差别，而答题

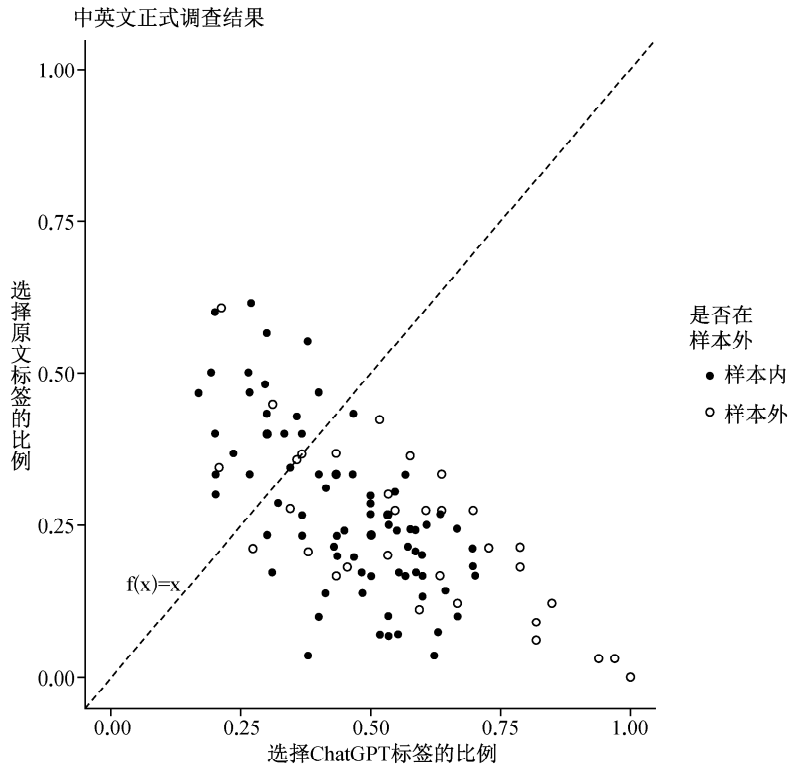


图 2　受访者选择原文标签与 ChatGPT 标签的结果对比

数据来源:Credamo 和 Mechanical Turk 正式调查数据。

人还是更倾向于选择 ChatGPT 标签(图 2)。这是为什么呢?

观察 ChatGPT 标签与原文标签的差别,我们提出下面的假说:与原文标签相比,ChatGPT 标签往往文字更多、更具体。这可能说明,用更多文字对主题进行更清楚的说明更容易说服读者。即便抛开本文采用大语言模型的方法探索不谈,这也给今后采用主题模型进行研究的学者以启发。作者给主题打什么标签本来没有任何限制,此时高度抽象概括的用语尽管看似言简意赅,实际上未必最有说服力。

表 2、表 3 分别列出剔除缺失值后,选择原文标签比例最高和最低的 10 个主题。其中表 3 报告的第 4 个主题就是前文提到的两个标签一致的主题,大多数人选择了"差不多",碰巧选择我们的底层代码指认为原文标签的答题人只有

一个。这两个表中有些结果确实符合前面的假说,但有些结果中 ChatGPT 的标签更为详尽,被选择的比例反而更低。对于这种情况,恐怕还需要进一步专门研究。

表 2　　所有论文中原文标签被选中比例最高的 10 个主题

编号	论文名	特征词	原文标签	ChatGPT 标签	原文占比
5	Managing the Boundaries of Taste	Adam、stroke、garag、pavement、pixi、surf、garagerock、weezer、lewi、malkmus	garage	Indie rock music	0.62
26	历史社会学能化解学科之争吗	women、family、children、men、marriage、household、work、social、female、gender、male、marry、live、age、class	性别与婚姻	性别和家庭动力学	0.61
4	Rhetorics of Radicalism	social、unity、political、communities、revolution、parties、west、exit、nature、system	social & political revolution	political transformation	0.60
3	Theming for Terror	faction、sharia、jawlani、Syrian、regime、god、Jesus、caliphate、camp、plot、Syria、pkk、pagan、shall、Ramadan	religious boundaries	Syrian conflict and extremist groups	0.57
1	Managing the Boundaries of Taste	Steven、bird、Stewart、muse、snow、finn、piano、stori、tale、narrat	lyrics/songwriter	storytelling or narration	0.55
8	Using Radical Environmentalist Texts to Uncover Network Structure and Network Features	action、direct、campaign、group、involv、network、sabotag、anti、reclaim、tactic、meet、success、event、sab、act、target、rts、opencast、media、aim	direct action/ecotage	activism and resistance	0.50
5	Seeing Like the Fed	credit、financial、institutions、risk、market、loans、banks、downside、cut、turmoil、losses、cdo、mortgage、deterioration、tranches、capital、auction、liquidity、housing、October、asset、January、sheet、swap、lending、commercial、agencies、senior、insurance、paragraph	financial markets	financial risk and instability	0.50

(续表)

编号	论文名	特征词	原文标签	ChatGPT标签	原文占比
3	Seeing Like the Fed	reserve、target、zero、program、funds、sheet、bound、rate、federal、December、quantitative、treasury、banks、facility、guarantee、interest、purchases、ceiling、quantity、effect、deflation、tools、excess、money、policy、fomc、size、monetary、desk、alternative	policy response	monetary policy tools and actions of the federal reserve	0.48
8	Analyzing Meaning in Big Data	company、year、percent、last、analysts、earnings、share、sales、years、quarter、profit、revenue、profits、analyst	stock market and profits	financial performance	0.47
3	Differentiating Language Usage through Topic Models	ident、practic、discours、cultur、nation、construct、global、wai、local、polit、examin、negoti、power、ethnograph、explor、argu、subject、relat、project、space	identity studies	cultural and political discourse in global and local contexts	0.47

表 3　　所有论文中原文标签被选中比例最低的 10 个主题

编号	论文名	特征词	原文标签	ChatGPT标签	原文占比
6	“黑箱”中的青年	推荐、喜欢、内容、隐私、方便、信息、感兴趣、符合、信息茧房、时间、推送、容易、娱乐、喜好、需求、侵犯、淘宝、音乐、商品、泄露、窥探、搜索、监视、有好有坏、中立、满足、兴趣、有利有弊、单一	娱乐内容	用户偏好和隐私	0.00
4	“黑箱”中的青年	类似、历史、浏览记录、监控、偷听、热门、屏蔽、偷窥、窃听、总是、反复	运作机制	监控和隐私担忧	0.03
8	“黑箱”中的青年	广告、不喜欢、隐私、讨厌、推荐、内容、朋友圈、算法、推送、感觉、侵犯、反感、投放、精准、很烦、太多、不感兴趣、泄露、监视、软件、平台、体验、烦人、消费、厌恶、厌烦、社交、浪费时间、用户、重复	广告内容	对广告推送和侵犯隐私的负面态度	0.03

（续表）

编号	论文名	特征词	原文标签	ChatGPT 标签	原文占比
6	Analyzing Meaning in Big Data	insurance、company、life、companies、policies、coverage、insurers、insurer、policy、percent、first、claims、premiums、business	Insurance industry	insurance industry	0.03
4	Racialized Discourse in Seattle Rental Ad Texts	winter、warm、morn、nearest、keep、climat	Cozy and Comfortable	winter weather and climate forecasts	0.03
19	历史社会学能化解学科之争吗	state、govern、new、public、policy、social、develop、nation、institution、politic、economic、city、universal、power、local	制度演变	公共政策与治理	0.06
3	Analyzing Meaning in Big Data	industry、congress、commission、new、government、federal、agency、law、legislation、rules、committee、administration、proposal、officials	Legislation	government and legislative affairs	0.07
6	Differentiating Language Usage through Topic Models	commun、member、organ、structur、network、interact、relationship、group、individu、relat、societi、kinship、form、activ、exchang、role、share、associ、maintain、institut	Social structures	social structure and organization in communities and groups	0.07
9	Racialized Discourse in Seattle Rental Ad Texts	cours、golf、Jefferson、height、point、sand	High-Class Surroundings	golf course geography and topography	0.07
6	Racialized Discourse in Seattle Rental Ad Texts	burk、trail、gilman、microsoft、campus、googl	Commuting Distance	technology companies and outdoor activities	0.07

当然，答题人的选择比例较高并不在客观上等价于这个标签更恰当。因为受访者未曾阅读原始语料，他们对标签的评判不具有专业性。例如，在格里夫等人论文（Greve，Rao & Vicinanza et al.，2022）的一个主题上，原文标签是“虚假死亡统计数字”，而 ChatGPT 标签是“死亡率和死因”。尽管选择 ChatGPT 标签

的答题人更多,但实际上原文是研究线上阴谋论的,所以原文标签才更准确。

这个问题并非无法解决,一旦我们在提示语中提供了正确的背景信息,ChatGPT 就会进行修正。我们在输入上述主题时还尝试给 ChatGPT 提供背景信息:“我们正在研究推特上关于 COVID-19 的阴谋论言论,并在这些阴谋论言论的语料中训练了一个主题模型。”ChatGPT 便返回了更加贴近研究问题的新标签:“COVID-19 Death Toll Comparisons”。这篇文章发表于 2022 年,在 ChatGPT 的知识阻断点之后,因此这个结果不是简单的信息检索得出的。这说明如果在提示语中给予充分的背景信息,ChatGPT 的编码效果还能进一步提升。前面本文运用的是不提供背景信息的相当简化的提示语,所得结果中 ChatGPT 的表现已经相当不错,因而如果改进提示语,想必在类似的评估中表现会更好,这里不必赘述。

这提示了大语言模型与主题模型在社会科学中的另一种结合方式。芒克(A. Munk)等指出,机器学习模型在预测上失败的案例,往往也是诠释起来最有难度、最需要研究者详细说明的案例(Munk, Olesen & Jacomy, 2022)。同理,也许 ChatGPT 标签与原文标签差异最大的主题,恰恰就是最富理论复杂性而值得探讨的主题。大语言模型此时从反向发挥了启示研究者的作用。

(四)局限

GPT-3.5 对于输入的信息有长度限制。由于要输入的内容较多,笔者很难将所有主题及其特征词一次性输入 ChatGPT,让它将不同主题之间的差异纳入考虑。出于同样的原因,某个主题占比最多的文章是哪篇、带有主题特征词的关键性段落包含什么内容、主题在时间维度上的比例变化等信息也没法被输入 ChatGPT,尽管这些信息在人为解读主题的时候是相当有价值的。

在这个问题上,GPT-4 有所改观。接下来,研究者可以评估这样的输入方式能否提升表现。可以预计,由于这些限制在工程上并非不可接近的,随着该技术路线得到认可,今后的大语言模型可能会进一步扩展其上下文的限制,并且能够接受文本以外的其他输入类型。尽管我们还不能对此进行评估,但基于现在的证据,我们对此表示谨慎乐观。

四、结论与讨论

本文结合大语言模型(ChatGPT)和小语言模型(主题模型),试图解决后者在应用中遇到的一个实际问题。由于给主题模型的结果打标签具有一定的任意性,而进行人工校验的成本又较高,因此对主题模型的结果进行大规模的标签校验难度较大。大语言模型基于其在预训练时所掌握的人类语言模式,可以自动给主题模型的结果打出相对而言可信度较高的标签,从而为上述问题提供一个可行的解决方案。

我们并不主张 ChatGPT 可以完全取代研究者的工作,不过它确实能够辅助研究者。首先,在模型探索阶段,大语言模型能够快速给模型的结果赋予标签,从而允许研究者在较短时间内探索更多的模型。其次,研究者可以将大语言模型标签与自己的标签进行对比,识别具有理论意义或需要澄清的主题。

基于这一发现,我们既对大语言模型模拟人工编码的能力抱有积极态度,又对实践中应用大语言模型持谨慎态度。在积极态度上,对于使用主题模型的论文而言,论文评审人进行匿名评审时的处境与本文中答题人的处境相似。他们很可能对这一研究使用的语料并不熟悉,只能依赖作者在论文中报告的特征词来对主题取得初步的理解,因而作者给主题赋予的标签就具有某种不容辩驳的权威性,而这恰恰可能关涉到后续的理论分析。由于作者在打标签问题上利益相关且可能存在偏误,评审人又可能缺乏纠偏能力,那么该如何保证人为标签不扭曲研究结果呢?大语言模型可能提供了一种检验的可能。但在消极态度上,评审人又能够根据什么判断原作者对语料确实有相当程度的掌握,而不是仅仅使用了大语言模型呢?在研究中禁止使用 ChatGPT 这类应用恐怕是不可行的,建立这样的规范只会让研究者欺瞒评审人。这一研究规范与伦理问题,可能是目前学术出版工作面临的最棘手、最紧迫问题之一。

这一研究还给了我们一些启发。如果研究不给主题模型的结果打标签,或者使用非常模糊、宽泛的标签,其结果并不容易说服读者。ChatGPT 给出的标签更具体、详尽,受一般读者的欢迎,说明研究者应该注意打标签的风格。在后续研究中,研究者可以进一步衡量:如果进一步扩展大语言模型的上下文限制,或

者使用小模型的训练文本对大模型进行微调,是否可以让大语言模型更贴近研究者的需要,取得更好的表现?

五、余论:“失控”

如果本文是一篇四平八稳的量化研究,那么到此本应已经结束。但从事研究的过程给了我们很多自反的机会,值得与学界同人探讨。

如果大语言模型已经可以很好地完成自然语言处理任务,那么为什么还要执着于主题模型这类小模型呢?正如文献综述部分介绍的,ChatGPT 等大语言模型对人类语言的理解比基于“词袋”的主题模型更复杂,对后者可能有替代性。用大语言模型改进小语言模型岂非刻舟求剑?

或许大语言模型尚未一统江湖的一个原因在于,研究者对小语言模型有更强的掌控能力。主题模型的原理并不太难,通过阅读主题的特征词而猜测其打标签的过程也颇直观。研究者可以在电脑上进行主题模型的运算,还可以调整超参数来探索模型的结果。尽管实际上主题模型内部蕴含着不可小觑的不确定性,但在一定程度上,研究者似乎还是可以把控住从输入数据到输出数据的过程。

相反,大语言模型内部结构的庞大与复杂性,已远远超出大多数研究者的认知能力范畴。在本文中,利用 ChatGPT 进行标签标注时,我们面临的一个主要问题是其可重复性无法得到保证。具体而言,由于模型本身固有的不确定性,即便是输入相同的提示语,每次得到的结果都可能存在差异。这首先是因为 GPT-3.5 如同其他概率模型一样,在推断过程中包含随机因素。除此之外,随着模型的持续迭代,即便是相同的提示语,在不同时间点的输入也可能产生不同的结果。这种结果的变化及其规律,完全超出了研究者的控制范围。至于为何某些提示语能够取得较好的效果,大多数人往往难以给出严谨的解释,这种现象仿佛带有某种巫术或魔法的色彩。

这明显违背了学界理想状态下的规范。理论上,研究工具的使用者应当对其工具的用途和局限有深刻的理解,并在必要时进行恰当的修正,从而对研究结果负责。然而,尽管滥用统计工具在学术研究中屡见不鲜,但在应然层面,这仍

然是一条被广泛认同的原则。然而,大语言模型这类高度复杂的研究工具,往往让使用者在多数情况下难以理解其原理、用途和局限,从而难以进行适当的评估和修正,进而无法对研究结果负责。与此同时,大语言模型在定义清晰的任务上表现出惊人的能力,使得传统的、相对落后的工具无法与之相提并论。

这种能力和责任上的对立,正是近年来计算社会科学中所遇到问题的缩影。新方法在诸多问题上的表现远胜于研究者过去所学的传统工具,但这些新方法往往也伴随着高度的复杂性,使得研究者难以充分掌握。在这种情况下,若仍坚持使用新方法,研究者则可能面临方法逐渐“失控”的问题。这里所说的“失控”并非一个绝对的状态,而是一个程度的变化,研究者可能从“不那么失控”逐渐走向“越来越失控”。

以 ChatGPT 为代表的大语言模型应用使得我们的方法“失控”达到了新的高度,这主要源于它与以往计算方法存在本质差异。过去的计算方法应用领域边界明确,我们往往对模型的数学估计过程有着较为清晰的理解(如 LDA 模型),或是模型针对特定任务而设计(如深度神经网络在分类预测任务中的应用)。这些方法伴随着一套专业术语,使得它们与社会科学的其他部分有所区分。因此,这些工具最终通常只能被部分人以相对规范的方式使用和理解,难以全面改变社会科学知识的生产方式。

然而,大语言模型则截然不同。一方面,其复杂性超出了绝大多数社会科学家所能掌控的范围;另一方面,其输出结果又极为接近自然语言,因此更容易被广大社会科学家所使用和理解。从理论上讲,大语言模型执行的是一项明确的预测任务,即通过对大量文本数据进行预训练后建立模型,以在给定的上文基础上预测下文。然而,由于这一任务的泛化性,研究者在使用时往往不局限于理解其预测结果,而是为其赋予了更深层次的意义。以本文为例,大语言模型实际上是在预测与提示后续可能出现的文本,而非理解、阐释或标注主题模型的主题,但我们却试图通过前者达到后者的效果。至于为何大语言模型能在这一具体任务上表现出色,我们仍感到困惑。这是因为训练模型所使用的目标函数并非研究者为特定研究任务所设定的,研究者对其细节既不了解也无法修改。有研究者称这种现象为大语言模型的“涌现”特质,但这一概念的提出似乎并未解决任何问题。

在大语言模型时代,机器学习模型领域正展现出日益封闭的特性,这进一步加剧了大语言模型难以被社会科学家掌控的困境。由于大语言模型规模庞大、投入巨大,已远超出个人乃至机构研究者的能力范畴,我们不得不依赖企业提供现成的产品,并在此基础上进行有限的调优。最初的大语言模型,如 BERT 和 GPT,均基于开源数据进行训练。然而,随着竞争压力的加剧和潜在法律风险的增加,开发大语言模型的公司已不再公开模型训练的语料来源和策略。以开发了 ChatGPT 的 OpenAI 为例,在公布下一代大语言模型 GPT-4 时,对于相关信息只字未提。这导致研究者难以了解企业使用何种数据进行模型训练,以及训练过程中可能存在的未披露问题。一旦企业产品升级换代,研究者就只能尝试新品,而无法确保此前研究结果的可延续和可推广。这些问题使得研究者对大语言模型的控制能力变得空前微弱,难以对研究结果承担高度责任。

与此同时,大语言模型的输入与输出却高度贴近自然语言,无须跨越专业术语的门槛。即使对统计方法、计算机语言一无所知的研究者,也能通过与 ChatGPT 的自然对话获得所需结果。ChatGPT 能够阐释线性回归表格的结果,推测访谈中话语的情感(尽管其可靠性有待商榷),甚至引经据典地写作文献综述(尽管引用的文献未必真实存在)。这种输出结果极难被人或机器分辨,意味着计算方法首次能够全面影响社会科学研究共同体中各个成员的生产方式。这或许能极大提升学术生产力,但也引发了人们的担忧:若我们对此习以为常,又如何分辨它提供的是知识、意见还是幻觉?

社会科学研究正面临方法“失控”的挑战,这对学界关于知识定义、知识生产方式以及知识产权归属的认知带来了巨大挑战。既往的规范中并未提供现成的答案。对于中国社会科学而言,要构建自主的知识体系,必须通过公开的研讨来解决这些问题。若处理得当,这或将成为中国社会学发展的一大机遇。

参考文献

陈云松、吴晓刚、胡安宁等,2020,《社会预测:基于机器学习的研究新范式》,《社会学研究》第 3 期。

郭台辉、周浥莽,2020,《历史社会学能化解学科之争吗?——基于西方学术史的结构主题模型分析》,《社会学研究》第 3 期。

赵龙轩、林聪,2022,《“黑箱”中的青年:大学生群体的算法意识、算法态度与算法操纵》,《中国青年研究》第 7 期。

周涛、高馨、罗家德,2022,《社会计算驱动的社会科学研究方法》,《社会学研究》第 5 期。

Bender, E., T. Gebru & A. McMillan-Major et al. 2021, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.

Blei, D. 2003, “Latent Dirichlet Allocation.” *The Journal of Machine Learning Research* 3.

Blei, D. 2012, “Probabilistic Topic Models.” *Communications of the ACM* 55(4).

Buurma, R. 2015, “The Fictionality of Topic Modeling: Machine Reading Anthony Trollope's Barsetshire Series.” *Big Data & Society* 2(2).

Caliskan, A., J. Bryson & A. Narayanan 2017, “Semantics Derived Automatically from Language Corpora Contain Human-Like Biases.” *Science* 356(6334).

Devlin, J.; M.-W. Chang & K. Lee et al. 2019, “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1.

DiMaggio, P., M. Nag & D. Blei 2013, “Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U. S. Government Arts Funding.” *Poetics* 41(6).

Goldenstein, J. & P. Poschmann 2019, “Analyzing Meaning in Big Data: Performing A Map Analysis Using Grammatical Parsing and Topic Modeling.” *Sociological Methodology* 49(1).

Greve, H., H. Rao & P. Vicinanza et al. 2022, “Online Conspiracy Groups: Micro-Bloggers, Bots, and Coronavirus Conspiracy Talk on Twitter.” *American Sociological Review* 87(6).

Grimmer, J. & B. Stewart 2013, “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21(3).

Heuser, R. & L. Le-Khac 2012, “A Quantitative Literary History of 2958 Nine-A Quantitative Literary History of 2958 Nineteenth-Century British Novels: The Semantic Cohort Method.” *Pamphlets of the Stanford Literary Lab* 4.

Kaplan, J., S. McCandlish & T. Henighan et al. 2020, “Scaling Laws for Neural Language Models.” arXiv: 2001.08361[cs.LG].

Kozlowski, A., M. Taddy & J. Evans 2019, "The Geometry of Culture: Analyzing the Meanings of Class Through Word Embeddings." *American Sociological Review* 84(5).

Li, L., L. Wu & E. James 2020, "Social Centralization and Semantic Collapse: Hyperbolic Embeddings of Networks and Text." *Poetics* 78.

Marshall, E. 2013, "Defining Population Problems: Using Topic Models for Cross-National Comparison of Disciplinary Development." *Poetics* 41(6).

Meeks, E. & S. Weingart 2012, "The Digital Humanities Contribution to Topic Modeling." *Journal of Digital Humanities* 2(1).

Miller, I. 2013, "Rebellion, Crime and Violence in Qing China, 1722: A Topic Modeling Approach." *Poetics* 41(6).

Mohr, J. &P. Bogdanov 2013, "Introduction Models: What They Are and Why They Matter." *Poetics* 41(6).

Munk, A., A. Olesen & M. Jacomy 2022, "The Thick Machine: Anthropological AI Between Explanation and Explication." *Big Data & Society* 9(1).

Nelson, L. 2021, "Cycles of Conflict, A Century of Continuity: The Impact of Persistent Place-Based Political Logics on Social Movement Strategy." *American Journal of Sociology* 127(1).

OpenAI 2022, "Introducing ChatGPT." https://openai.com/blog/chatgpt.

Ouyang, L., J. Wu & J. Xu et al. 2022, "Training Language Models to Follow Instructions with Human Feedback." arXiv: 2203.02155[cs.CL].

Prepare for Truly Useful Large Language Models 2023, *Nature Biomedical Engineering* 7(2).

Stoltz, D. & M. Taylor 2021, "Cultural Cartography with Word Embeddings." *Poetics* 88.

Ying, L., J. Montgomery & B. Stewart 2021, "Topics, Concepts, and Measurement: A Crowdsourced Procedure for Validating Topics as Measures." *Political Analysis* (September).

数字时代的科技公司与国家治理

——一个综合性的分析框架

钟瑞雪*

摘要：在数字时代，科技公司与国家治理之间的关系属性和互动机制呈现出一幅跨时代的新图景，科技公司在国家治理中扮演着愈发重要和复杂的角色。这一新变化引起了学界的广泛关注，学界从不同角度对科技公司与国家治理的关系展开了研究。基于对既有文献的系统梳理，本文尝试构建一个综合性的分析框架，从“治理主体—治理工具—治理对象”这三个维度来考察数字时代科技公司与国家治理的关系。研究发现，数字时代的科技公司同时扮演着治理主体、治理工具和治理对象三种角色，而且三种角色密切关联、彼此形塑、共同演进，深刻影响着经济、社会的运行，使国家治理既充满革新又高度复杂。这一探讨不仅有助于深化对科技公司在国家治理中角色的理解，也为推进数字社会的研究提供了可能的分析思路。

关键词：科技公司　国家治理　治理主体　治理工具　治理对象

一、引言

进入数字时代，国家治理体系正发生着深刻的变革，其表现之一就是科技公司与国家治理的关系日益密切且复杂。科技公司所引领的信息技术革命正拓展着国家治理的新领域（徐汉明、张新平，2018），不仅使国家治理的对象发生了巨大变化，其自身在国家治理中所扮演的角色也发生了显著变化，这些变化引起了现有研究的广泛关注。比如，已有研究发现，科技公司掌握着数字技术与平台数据，基于这种技术与数据的新型技术型权力正在生成，并加剧着权力失衡（梅立

* 钟瑞雪，中国社会科学院大学社会与民族学院。

润,2023a;黄其松,2022;Andrejevic, 2014);又如,在各层级政府的治理创新中都可以看到购买科技公司服务的现象,科技公司成为国家治理的重要支撑力量(陈涛、董艳哲、马亮等,2016;王伟域,2022);再如,人工智能的应用与平台经济的兴起给国家治理带来了新的挑战,是国家治理面临的新对象、新议题(高奇琦,2020;周文、韩文龙,2021;孙晋,2021)。

根据经济合作与发展组织(OECD)的分类标准,高技术包括电子与信息技术、生物工程和新医药技术、新材料及应用技术、先进制造技术等。本文中的"科技公司"是指依赖于技术创新,包括电子科技集团、互联网企业、通信科技公司等企业在内的配备数字化硬件设施(杨志云,2019)、构建复杂经济网络(江小涓、黄颖轩,2021)、加速现代信息交流(乔天宇、邱泽奇,2020)等,以研发、生产和销售高科技产品与服务为主要业务的企业。在前数字时代,科技公司相对较少,或者可以说尚未充分涌现。20 世纪 90 年代以来,科技公司迅速崛起并广泛渗透到各行各业,英美等发达国家及中国先后进入数字治理时代,这为科技公司与国家治理间的相互作用奠定了基础。在整体治理体系中,科技公司与国家治理间的关系可能会产生一些增量变化,甚至质的变化。

首先,从企业自主权的演变来看,在前数字时代,我国国有企业(周飞舟,2019)与私营企业(史晋川、钱滔,2005)的发展,都经历了与当地政府和国家的依赖关系逐渐衰弱,并且逐渐掌握国家所赋予的自主权的过程。其次,从企业技术的应用场景来看,在前数字时代,企业的技术应用主要集中在企业内部,且聚焦于生产领域,如进行企业内部绩效提升、组织模式选择等调适(吴晓波、陈颖,2014;胡建绩、孔德洋,2000)。科技公司与国家治理的关系还没有如此密切,通过某种技术来为国家治理提供帮助,甚或将企业技术直接运用于国家治理还比较少见。在通信不便、信息传递存在较大障碍的情境下,企业和国家相对封闭,技术创新的重点主要集中在组织内部的发展,较少涉及全局层面的协同合作。最后,从组织的影响范围与程度来看,企业的社会影响被严格划定在组织和产业内部,或发生于平行的企业组织之间,具有明晰的边界。国家是处理公共事务的主体,其治理的开展主要依靠科层组织内部上下级政府之间的互动(周雪光,2008)。周黎安也指出,我国政府间关系更像是组织内部层层发包的关系(周黎安,2017)。这时的企业与国家所产生的影响均局限在组织内部。国家在很大

程度上控制着绝大多数重要的社会资源，企业主倾向于利用社会关系来从国家手中获取相应的社会资源（王晓燕，2006；黄波，2010）。

但进入数字时代以来，上述三个方面都发生了重大变化。首先，科技公司的自主权产生了新的变化。这里的“自主权”不仅涉及在科技公司协助国家治理的过程中政府所让渡的话语权等权力，还囊括其基于自身技术与数据优势所获得的技术垄断、数字独裁等权力（叶娟丽、徐琴，2019）。比如，梅立润指出，在数字时代，数据和算法是新的权力生产要素，掌握着数据和算法的科技公司或互联网企业便掌握着技术型权力（梅立润，2023a）。其次，科技公司的数据资源多、技术水平高，能够衍生出适用于商业应用场景之外的“看门人”等功能（单勇，2023），因而对国家而言具有了超越产业本身的价值。国家治理从未像今天这样需要依靠科技公司的力量（向静林、钟瑞雪，2024），国家也从未像今天这样需要面对数字空间中的治理对象，而数字空间的重要载体便是掌控着数字平台的科技公司。最后，在数字时代，每个人都成为科技公司所创制出的平台上的成员，平台和网络成为人们生活的基本场域，形塑着人们的日常生活和所思所想（金兼斌，2019），其所包含的信息、人群、资金等内容在全社会范围内产生了广泛影响，传统社会正颠覆性地成为数字社会或平台社会。以金融平台的发展为例，当网络所具备的连通性特征使得金融交易不再受时间、空间和区域的限制时，平台的用户数量和使用范围在短时间内急速扩张，对国家治理的快速响应提出了挑战（向静林，2021）。上述特征的更动表征着科技公司与国家治理的关系出现了划时代的变化，呈现出一幅新型的国家和企业关系图景。

但目前，上述的研究工作是碎片化的、比较分散的，未能建构起探讨数字时代背景下科技公司与国家治理间关系的整体性和系统性分析框架。对这一问题进行梳理和深挖，有助于推动这一领域的研究，提供综合性的理论视野。那么，数字时代科技公司与国家治理的关系究竟如何，这种新型关系又如何从理论上加以理解？这是本文的研究问题。本文在已有文献的基础上，尝试总结出“治理主体—治理工具—治理对象”的三维分析框架，展现科技公司与国家治理的复杂性关系。

二、作为治理主体的科技公司

科技公司掌握基于数字技术和数据的权力，这种权力不是由其他主体授予的，而是源于它在经济社会中的不可或缺性，甚至是垄断性。这种权力体现在很多公共事务的治理实践中，如科技公司通过甄别网络谣言，利用平台“主管”的身份和掌握传播趋势的优势构建出应对网络谣言的安全管理体系（金璇、赵玉现，2017）；在线上违法犯罪治理中，科技公司作为平台的载体，能够及时监测到违规行为并采取有效处罚，从而获得犯罪治理的权力（单勇，2020）；在轨道交通治理中，科技公司能够利用大数据提前掌握交通流动情况，建立交通拥堵预警和防治系统，降低交通堵塞的发生率，为交通管理减轻负担（周辉宇，2017）；等等。

在数字时代，科技公司以平台为突破口，强势地贯穿了人类社会生活的方方面面，其自身所具备的开放式、创新性、技术导向等特征（Chesbrough，2003；Tushman & O’Reilly，1996；陈鹏，2019），以及研发的线上会议室、社交平台和移动支付平台等产品正持续改变和塑造着用户的工作、沟通和交易方式。比如，互联网金融模式不仅提高了金融普惠性，还通过降低市场不对称程度等形式大大降低了交易成本（谢平、邹传伟，2012）。不仅如此，科技公司还能够协助政府解决社会问题，推动社会公平。在以往的政府实践中，各地环保部门对于污染企业偷排现象只能采取事后处罚制度。2017 年 6 月，阿里巴巴旗下的阿里云启动了 ET 环境大脑项目，与江苏等多地环保部门开展项目合作，帮助政府对污染源进行全方位监测，在事发前进行专业预警和识别（关婷、薛澜、赵静，2019）。

作为诸多平台或技术下的受益群体，用户也为科技公司的生存与发展奠定了合法性基础（张茂元，2021）。不断提供新技术产品和新型组织形式满足社会需要的发展路径，为平台和科技公司获取用户的行为数据创造了可能。然而，数据本身没有价值，因其经过技术收集和加工后为平台创造流量、带来资源而具有了价值。一方面，平台不断将用户的现实生活转化成数字符号，使生活普遍数据化（解正山，2020）；另一方面，平台基于用户“数据化”和大量的用户基础获得了海量的数据资产。人们在接触技术产品的过程中不断贡献着数据，成为“数据人”。比如，各类数字平台上的实名认证机制最初是为确定用户真实身份、保证

用户安全而设立的,但现实却是为科技公司收集并储存用户的基本信息提供了便利,而这些信息可以被用来换取利益(金兼斌,2019),进行资本再生产。

除却个体的日常生活,随着集社交、交易、信息共享等多功能于一体的数字平台的兴起,组织的运营行为也逐渐被纳入其中并被数据化。平台不仅掌握着大量的数据以及流量,还通过聚集、整合、分析等手段使数据成长为巨大的数字生态,各大平台几乎以垄断的方式占据着这些数字生态,使其他的用户不得不购买这些科技公司的服务(蓝江,2022),并借此机会俘获这些合作用户的数据。同样,当政府借助平台处理政务或将部分政府事务外包给科技公司处理时,政府内部的大量数据也在向企业透明化(叶娟丽、徐琴,2019)。科技公司作为平台的搭建者与实际运营者,可以充分利用平台的门槛机制、监督惩戒机制对平台内组织和用户的行为进行监督审查(阳镇,2018),实现有效治理。从宏观的视角出发,公众也更倾向于使用不同的平台来与政府部门协调办理相关事务(宋锴业,2020)。国家作为社会治理的主要力量,所面临的治理对象由具体的现实场域转变为虚实融合的数字空间,对数据的需求随之提高(梅立润,2023b)。因此,作为数字空间的创造者和海量数据信息与技术的掌握者,科技公司无疑成为国家治理数字社会的一大行动主体。

伴随着科技公司对人类生活的基础性渗透,国家及公众对科技公司产生的资源依赖性成为其技术权力的源泉(金兼斌,2019),掌握着数据和算法的科技公司即拥有这种技术型权力。同时,科技公司对数据有着绝对的垄断型权力(Danah & Crawford, 2012),且这一垄断会在数字时代持续加深(高奇琦,2020)。安塞尔(C. Ansell)和缪拉(S. Miura)指出,科技公司从根本上构建了一套新的组织逻辑(Ansell & Miura, 2019):一方面,政府在利用平台的数字技术对自身功能进行调试与升级的同时,也对自身的运作模式进行了更新;另一方面,科技公司在合作的过程中借助政府的科层组织体系形成了符合自身特征的组织属性(Ansell & Gash, 2018),并改变着政府与社会的互动关系(宋锴业,2020)。表面上,国家权力独立解决了大部分社会问题,但也有部分社会问题的解决需要科技公司的技术协助,而正是这一过程使科技公司获得了国家委托或授予的权力。此外,部分科技公司还会通过游说渠道来影响国家决策和舆论导向,使公司自身的规则成为行业准则或法律法规(Huang & Chen, 2020),使平台

能够依托自身所具有的技术与市场环境，加速相关垄断市场的形成，形成一批集技术、资本、权力为一体的超级平台（占佳，2021）。更为甚者，通过还原或构建类似真实世界的组织与社会，构建新型的人际交往关系并连通真实世界的万物，平台从一个只具备服务性质的场地发展成一个虚拟世界。最终，科技公司将裹挟着处于真实世界的个人进入这一超级平台，实现以虚拟世界取代现实世界的目的（赵汀阳，2022）。运营数字平台的科技公司会逐渐发展成一股新的超级权力，成为实际掌控社会权力的主导性力量（单勇，2023），在对数据的掌控与管理等层面挑战国家的权力地位，引发严重的权力失衡危机（Andrejevic，2014）。

党的十八届三中全会公报指出："全面深化改革的总目标是完善和发展中国特色社会主义制度，推进国家治理体系和治理能力现代化。"党的二十大报告进一步强调2035年基本实现国家治理体系和治理能力现代化的总体目标，快速响应数字时代国家对治理现代化的号召。在新的国家治理图景中，尤为突出的是社会治理主体的多元化，其中便包括掌握现代技术的科技公司。但是，需要清楚的是，能够参与到社会治理中的科技公司是获得国家允许的。因此，在整体治理过程中，国家仍处于主导地位，企业等其他治理主体所获取的国家让渡的权力非常有限（王锋，2022）。而且，在传统社会向数字社会转型的过程中，会打破固有的社会秩序，新秩序的产生也伴随着利益重新分配等一系列问题，此时需要国家发挥稳定器功能，行使国家权力，将人作为治理的中心，保护国民的生命财产安全，保证社会正义（高奇琦，2020）。

值得一提的是，科技公司在国家治理中的主体性角色和情境有关。在不同领域、不同层级以及不同时期，科技公司的主体性存在一定差异。比如，新冠疫情到来之前，大多数平台并未将目光投向公共卫生领域，合作对象也多为各地方政府和企业，偏重于打造社区共建共治共享平台（陈荣卓、肖丹丹，2015）、推动中小企业数字化转型（李凌浩、吕鹏、刘金龙，2022）、深化互联网金融治理（艾云、向静林，2021）等。随着新冠疫情突袭而至，平台着眼于自身已有的产品优势在公共卫生领域进行优化，通过国家自上而下的要求和企业自下而上的匹配这两条轨道促成与更高层级政府部门合作的机会，由企业内部治理转向国家公共治理（吕鹏、周旅军、范晓光，2022）。因此，本文将科技公司在国家治理中的主体性看作一个变量，而不是一个常量。

科技公司通过其垄断性的数字技术和数据资源,积极地介入公共事务治理,塑造了一种新的权力形态。它们在各领域发挥着关键作用,不仅为政府提供了技术支持,还在数字经济中垄断了关键资源。这种数字权力使科技公司成为国家治理的有力主体,引发了国家和公众对其技术资源的依赖。不过,科技公司的主体性在不同领域和时期可能有所不同,其角色是一个动态的变量,受到国家政策和社会需求的影响。因此,在这个新的国家治理图景中,国家仍然占据主导地位,只是需要与企业等其他治理主体协调,以确保社会的稳定和正义。

三、作为治理工具的科技公司

当前,科技公司凭借大数据和先进技术渐进式地获取了一部分权力。这种权力不仅局限于商业上的垄断性和竞争力,还是国家治理现代化所需要倚赖的重要力量。党的十八届三中全会通过的《中共中央关于全面深化改革若干重大问题的决定》明确指出:“推广政府购买服务,凡属事务性管理服务,原则上都要引入竞争机制,通过合同、委托等方式向社会购买。”(新华社,2013)科技公司参与政府项目,开展政企合作,通过数字化技术和数据分析为政府提供了全面、高效的决策依据(北京大学课题组、黄璜,2020)。同时,它们参与到打击网络犯罪等领域,为社会问题的解决贡献力量。这种密切合作不仅提高了国家治理水平,也推动了社会各领域的创新和进步,科技公司与国家治理之间形成了相互依存的关系,共同推动着社会进步和治理现代化的发展。

政府购买公共科技服务,主要以合同委托、项目资助、经费补贴这三种方式实现。其中,合同委托是最为常见的方式,它制定明确的合同并规范双方责任,以确保政府购买服务的内容和要求得以明晰(杨燕英、胡文辉、田开学,2019)。随着政府购买服务的支出在财政预算中占比逐渐增加,政府购买服务的绩效问题备受关注。政府通过签订合同的方式将公共服务供给任务委托给科技公司,科技公司作为代理人按照合同规定提供公共服务,政府会根据科技公司履行合同的表现来支付相应费用(李洪佳,2020)。国际上,这种按绩效付酬的形式已经成为政府购买服务领域重要的治理工具,促进了绩效管理的实施(曹堂哲、魏玉梅,2019)。

科技公司承接政府项目，为提高国家治理水平提供技术支持。通过数字化技术，科技公司能够为数字治理领域的问题提供解决方案，并帮助政府更高效地管理公共事务、提供服务，如电子政务、智能城市建设等，这是政企合作中科技公司的主体任务（单勇，2019）。

首先，科技公司推动政府政务平台建设，提供了政府与公众互动的新途径。2015 年，习近平总书记首次提出“数字中国”这一概念。党的十九届五中全会强调加快数字化发展，坚定不移建设数字中国，并对数字政府建设做出系统布局。各地政府纷纷建设具有代表性的数字政府。全国现已有 30 多个省级政府构建了线上政务服务平台，其中多见科技公司的身影。比如，北京市 12345 市民热线为北京市有关部门和中国联合网络通信有限公司联合设立，该热线被广泛应用，通过收集反馈信息为政府决策提供了重要的数据支撑，显著提升了国家治理水平。浙江“浙里办”APP 由阿里巴巴等集团共同出资成立，这一平台包含 500 多项服务，实现了用“数据跑腿”代替“群众跑腿”，从“最多跑一次”到“根本不用跑”的目的。广东“粤省事”APP 由中国电子、腾讯等公司共同投资成立，将过去“以部门为中心”的办事模式转变为“以群众为中心”的办事模式，用户还可以在该平台上关联电子证照（姜红德，2020）。这些平台的建设不仅提高了政府效能，为公众提供了更便捷的政府服务，还为传统电子政务创造了创新升级的机会（Janssen & Estevez, 2013）。

其次，科技公司协助国家发展全新治理模式，使全面实现国家治理现代化成为可能。2015 年国务院印发的《促进大数据发展行动纲要》明确指出，国家治理须“建立‘用数据说话、用数据决策、用数据管理、用数据创新’的管理机制，实现基于数据的科学决策”（国务院，2015）。政府与企业，尤其是与科技公司的合作能够有效帮助其通过社交媒体、在线调查平台等软件了解民意、舆论动态，为政策调整提供数据支持（Jungherr，2016；孟天广，2021）。“为村”是腾讯公司于 2015 年 8 月面向中国广大乡村及社区推出的智慧乡村信息服务平台，它利用“互联网+”技术手段，搭建政商、政社、商社的多方协同沟通平台。通过设置网上规则，激励贫困地区精英力量积极争取、主动作为，再造互联网时代乡村精英，实现基层社会治理（吕鹏、房莉杰，2020）。“杭州城市大脑”整合交通、食品安全、法务等各类资源与数据，提供智能城市解决方案，包括智能交通管理、环境监

测、城市规划等,帮助政府提升城市管理水平,为多元主体协同治理提供支持(郁建兴、樊靓,2022)。2020 年 3 月,习近平总书记对“杭州城市大脑”运营工作与成果予以肯定:“让城市更聪明一些、更智慧一些,是推动城市治理体系和治理能力现代化的必由之路。”

最后,科技公司深入参与针对网络平台的治理,为营造绿色网络空间增添力量。以对网络平台犯罪的治理为例,数字技术与数字平台的涌现在提供多样化便捷服务的同时也助长了网络犯罪的发展(单勇,2023),而平台作为网络世界的看门人,为国家传统事后回应模式的危机提供了转变的契机(单勇,2022)。而且,当前网络犯罪治理的重心已从个人转向集团和组织(艾云、李祥、向静林,2023),科技公司作为平台的直接运营者,能够协助警方获取和保留线上犯罪证据,并根据数据信息进行犯罪过程追踪,对犯罪集团的生态链进行有效阻断(雷鑫洪,2021)。例如,网络诈骗使用第三方支付平台而不是银行等实体机构进行转账,为犯罪行为提供了更加隐蔽和安全的藏匿之地,而线上支付平台背后的科技公司却能够轻而易举地监管、识别这些诈骗行为。近些年,伴随着《关于加强打击治理电信网络诈骗违法犯罪工作的意见》《关于依法惩治网络暴力违法犯罪的指导意见》等一系列政策文件的出台,多家科技公司针对此类网络犯罪行为展开治理:在国务院联席办的指导下,百度发布“光明行动计划”,对于暴露用户隐私、散播虚假信息、买卖公民信息等黑灰产进行严厉打击,确保网络空间的清朗;腾讯专门成立“腾讯安全”,保卫用户、企业、政府三方网络安全,并推出腾讯云举报平台,为现实犯罪抓获提供便利(高海涛,2021)。

政府购买作为一种支持公共科技服务发展的新形式,也是科技公司参与国家治理的主要形式。它在实践中存在多重组织目标且相互关联,共同构成了综合性的目标体系。首先,最重要的目标是提供高质量公共服务以满足社会需求。平台能够通过意见的采集和整理来精准预测大众需求(容志,2019),并根据招标等竞争性购买形式的市场竞争提高服务效率(周俊,2014),推动经济增长的同时确保服务的社会包容性,保持合法合规的采购流程。其次,绩效管理的模式确保服务提供商履行合同,提高政府的治理能力,也促进公共利益的实现(朱火弟、蒲勇健,2003)。这些目标相互交织,共同为政府购买科技公司服务的成功实施打下坚实的基础。政府购买科技公司的服务,既能够促进国家治理向数字

化方向转变，提升公共财政资金的使用效率和政府的决策效率，又能够充分发挥企业效力在公共科技服务领域的重要作用，提高公共科技服务质量，是政府在公共科技服务领域重要的治理创新工具（杨燕英、胡文辉、田开学，2019）。

然而，由于我国数字技术发展与政府购买科技服务起步较晚，数字技术带来了一系列治理挑战（鲍静、范梓腾、贾开，2020），在公共服务层面的应用有待提升。首先，之所以衍生出政府购买科技服务的需求，主要是因为政府部门大数据人才与产品的缺失，国家治理现代化转型不得不协同新的社会主体。面对政府越来越频繁且广泛地依赖数字治理的现状，政府部门购买的公共服务在发挥有效价值、挖掘数据价值等方面，也因大数据人才的缺乏而存在掣肘。这督促着政府加速构建人才供给体系，提高人才储备规模（华子岩，2023）。其次，2018 年财政部发布的《政府购买服务管理办法（征求意见稿）》提出："购买主体选择项目承接主体时，要提高大数据运用能力，通过有关平台查询并使用相关服务供应机构的信用信息和信用报告，将其信用状况作为确定承接主体的重要参考。"（财政部，2018）但目前我国施行的法律法规不能实现购买流程的全覆盖，加之科技公司在数字技术上所占据的优势地位，购买服务后的管理与评估仍存在缺乏制度规范和被忽视的问题（徐家良、赵挺，2013）。最后，政府若期望开展基于数据的治理，那么政府各部门之间应相互配合，进行数据协同共享与整合，一方面需要避免目前"数据碎片化"的情况持续存在，另一方面也应该积极改变将大部分数据交与科技公司进行治理的局面，以及可能造成的影响政府购买公共服务精准性的结果（吉鹏、许开轶，2020）。

科技公司在国家治理中扮演着日益重要的角色，发挥着积极作用。它们通过政府购买服务等方式参与到政府的社会治理项目中，通过数字技术和数据支持，在公共事务治理、社会问题解决、政府运作升级等方面提供关键帮助，有助于提高政府的工作效率和决策质量，是推动国家治理现代化和数字化转型的有效工具。

四、作为治理对象的科技公司

在数字时代，科技公司凭借自身所掌握的数据、数字技术以及技术产品，不

仅广泛影响着公共事务治理,还改变着人类社会的方方面面,更成为国家治理所依赖的重要工具,助力社会治理数字化转型。这种对数据及技术的需求赋予科技公司以重要的地位,使科技公司获取日益增强的数据垄断权力,但也在一定程度上挑战了国家的地位。其庞大的数据储备和算法能力可能引发隐私泄露、信息滥用等问题,对公众权益构成潜在威胁,为社会治理带来风险(吕鹏、房莉杰,2020)。此外,科技公司的垄断地位和信息掌握能力使其在社会治理层面产生深远的影响,可能导致信息控制(杨明,2022)、市场不公平竞争以及破坏公众信心(冯乾、王海军,2017)等问题。因此,科技公司的迅速壮大不仅带来了便利,也引发了国家及公众对数据控制、权力失衡以及社会风险的担忧,需要在国家治理框架内进行监管和治理,以确保其发展不损害国家利益和公众权益,保证社会的稳定和公平。

如今,国家越来越意识到数字平台所带来的潜在威胁。国家互联网信息办公室发布的《数字中国发展报告(2021 年)》显示,数字平台上的虚假信息、网络诈骗、恶意软件等问题已经引起国家对网络安全的关注与应对。这些问题不仅会损害个人和组织的利益,还会对国家的整体安全构成风险(国家互联网信息办公室,2022)。同时,平台化趋势带来了治理场景的多元化,导致信息碎片化和多样化,催生出社会结构中的不公平和不确定现象(孟天广、赵娟,2018)。这些平台放大了潜在风险的影响范围,使得“社会风险容易向‘多领域’转移、向多维度延伸、向多平台扩散”(谢新洲、石林,2020),从而增加了实际危害和治理难度,挑战了传统的政府监管模式(樊鹏等,2020)。

具体来讲,在政治领域,国家主权及监管体系受到冲击。在数字技能短缺的情况下,国家数字主权可能会被减弱,从而降低国家对整体目标的掌控能力。特别是当科技公司具备较高的自主权时,它们更加倾向于以私利为导向,将国家为促进治理而向科技公司提供的公共资源和信息纳入私人或跨国领域,从而增加国家安全的潜在风险(高奇琦,2023)。然而,相比于科技创新的迅速迭代,社会适应却需要较长时间,这使得政府原有的单一类型的监管工具不能全面和适配地应对风险(魏小雨,2019)。

在经济领域,平台垄断问题在数字经济中较为突出。头部平台企业所采用的合并策略进一步巩固了其市场主导地位,对细分领域的竞争造成冲击,同时也

加剧了其滥用市场支配地位的动机。这种趋势不仅抑制了中小型企业的创新活力,而且对消费者和上下游企业的利益造成了损害(孙晋,2021)。比如,寡头数字平台通过长期积累的大规模用户数据和强大的数据管理及分析能力,拥有引领市场和巩固市场领导地位的能力,所引发的歧视定价等问题对经济和网络安全会产生潜在影响(熊鸿儒,2019)。

在社会领域,数字平台已经成为社会互动和信息传播的关键媒介。社交媒体平台汇聚了数十亿的用户,成为人们分享生活、沟通交流的主要工具(李明,2015)。然而,这些平台也容易成为不良信息传播的渠道和社会事件的发酵地(许晓龙,2015)。举例来说,互联网平台可能强化了不同观点之间的分歧,阻碍有效共识的形成,甚至削弱国家治理现代化的大众支持。此外,线下问题在线上环境中的扩散不仅使社会秩序和稳定问题更为复杂,还对其构成了严重威胁(谢新洲、石林,2021)。

可见,数字平台的崛起已经引发了一系列问题和挑战,国家基于对安全、公平、稳定等要素的关切,衍生出了针对平台进行治理的需求及机制。习近平总书记指出,我国平台经济发展正处在关键时期,要推动平台经济规范健康持续发展。2022 年,国家发改委等九部门联合印发了《关于推动平台经济规范健康持续发展的若干意见》,针对平台竞争监管、平台数据使用、平台价格行为和平台金融监管四个方面分别建章立制,突出强调国家的监管与治理职能。一方面,国家制定和修订相关法律法规,以规范数字平台的运营行为(熊鸿儒,2019)。这些法律法规包括反垄断法、数据隐私法、信息安全法等,以确保数字平台不滥用市场支配地位,维护市场竞争的公平性,保证平台运营的合规性。例如,《中华人民共和国反垄断法》第一章第九条规定:“经营者不得利用数据和算法、技术、资本优势以及平台规则等从事本法禁止的垄断行为。”《中华人民共和国个人信息保护法》第一章第十条规定:“任何组织、个人不得非法收集、使用、加工、传输他人个人信息,不得非法买卖、提供或者公开他人个人信息;不得从事危害国家安全、公共利益的个人信息处理活动。”《中华人民共和国数据安全法》第一章第八条规定:“开展数据处理活动……不得危害国家安全、公共利益,不得损害个人、组织的合法权益。”另一方面,政府采用合作监管中的超级综合监管模式(张丙宣、华逸婕,2019),负责监督数字平台的合规运营。举例来说,中国在数字货

币领域采取了严格的监管政策。从 2013 年比特币在中国允许交易，到 2021 年中国政府联合部委、银行、银保监会以及支付机构全面禁止所有形式的加密货币活动（陈伟光、明元鹏，2021），这一举措确保了数字货币在中国受到了有序的监管和禁止，以维护金融稳定和安全。此外，国家还会与其他国家寻求合作，共同应对跨国数字平台的挑战（陈伟光、钟列炀、裴丹，2022），这种国际合作有助于协调跨国问题和制定全球标准。

在政府对科技公司及平台进行治理的过程中，也催生出新型的政企关系，即头部平台企业借助其巨大的规模和资源，逐渐形成专门与政府打交道的事务部门（黄冬娅、杜楠楠，2022）。在传统政企关系研究中，企业主主要通过入党、在企业内部设立党组织等举措实现政治参与（朱光磊、杨立武，2004），或是通过非制度化的社会关系（王晓燕，2006；黄波，2010）来获取社会资源。在本文所提到的新型政企关系中，科技公司不再依靠个人化和传统化的政企关系，而是开设专项的政府事务部门，如公共政策部门、政务业务部、企业社会责任部等来达成上述目标。这些部门通常由经验丰富的公司高位职员组成，他们了解政府的政策、法规和公司内部的发展策略，而且这些人员与政府各级机构建立了紧密的联系，以确保平台运营符合国家法规与期望，同时也协助平台应对政策变化和新的监管要求（黄冬娅、杜楠楠，2022）。这一现象的出现不仅为政府提供了了解数字平台运营情况的更直接的渠道，还为平台企业提供了更好的机会来影响政策的制定。

然而，我国政府若要实现平台治理的良好效果，还需要进行长期探索与坚持。首先，政府的治理思路应该从以“先发展再治理”为导向，转向以“先构建治理框架再实现发展”为导向，重新定位政府监管的核心，让使用者承担更多的责任。这将有助于全面重构科技公司的治理结构（王锋，2022），以更好地满足科技公司复杂多变的需求。或者，针对数据运用行为展开治理时，可以采纳底线思维，这不仅允许科技公司在创新数据运用的范围和方式方面有更高的自由度，同时也强调创新不能损害利益相关方的权益，更不能损害社会的公平和公正（邱泽奇、乔天宇，2021）。其次，政府应改善监管理念与规则，发展与平台特征和动态发展相匹配的积极包容审慎监管、公平性监管、中央整体监管和大部制协同监管、自主性激励性监管、信用监管和智慧监管等监管模式（孙晋，2021），并在落

实中对平台进行分类监管，如搜索性平台、电子商务平台、社交类平台等（江小涓、黄颖轩，2021）。最后，从国家层面增强对科技公司使用数据、技术等信息的监管与法制化程度，同时鼓励平台自身规定平台治理的规范化，并发挥行政主管部门、第三方治理机构和社会公众的外部监督作用（孙晋、蔡倩梦，2023）。

在数字时代，科技公司的崛起对国家治理和社会产生了广泛而深远的影响，引发了隐私安全、信息滥用、社会稳定等问题。持续治理数字平台对于维护国家安全、促进市场公平竞争、保护个人权利和解决社会问题都具有重要的意义。这不仅能够实现数字时代的可持续发展，还能确保数字平台对社会产生积极的影响。政府和科技公司需要共同努力，以平衡数字平台的发展和公众权益，确保社会的稳定和公平。

综上所述，在数字时代，科技公司以其垄断性的数字技术和数据资源积极参与国家治理，为政府提供技术支持，推动国家治理的数字化转型。尽管与科技公司的合作提高了政府的工作效率和决策质量，但也引发了权力失衡、数据垄断和隐私关切等问题。科技公司积极参与各领域的公共事务治理，为政府提供技术支持，塑造了新的数字权力形态，推动了国家治理的现代化和数字化转型。然而，它也带来了一系列挑战，包括数据与技术垄断、信息滥用、社会稳定、公平竞争等问题。因此，相关领域的学者强调政府须积极采取监管措施以确保数字平台的合法性、公平性、透明性，倡导新型政企关系，鼓励科技公司规范自身治理与外部监督，以实现数字时代治理的可持续发展与公众权益维护。

五、结论与讨论

在数字时代，科技公司积极参与国家治理，提供数字技术支持，推动数字化转型，在提高政府效率的同时，也带来了诸多挑战。已有研究进行了深刻的洞察，意识到了科技公司在数字时代的重要作用。但是这些研究多零散地分布于各个领域，缺乏一个整体的、清晰的分析框架将其整合起来。在此基础上，本文提供了“治理主体—治理工具—治理对象”的三维分析框架，采用多维度的方式来审视科技公司在国家治理中的角色，打开科技公司三种角色间复杂关系的多重面向。从中我们能够看到每一种角色的特征，以及在每一种角色之下科技公

司与国家治理间关系的核心特征和差别。

从表 1 中我们可以清晰地看到科技公司三种角色的形成原因、特征、具体表现形式、结果及已有研究议题。这种划分为我们理解科技公司与国家治理的关系提供了一种清晰的思路,有利于展开看似混沌复杂的现象的内部结构。表 1 不仅关注了科技公司与国家之间的关系,还考虑了它们在治理体系中的地位,有助于识别和分析科技公司在国家治理中的不同角色,揭示科技公司如何影响国家权力、国家如何与之互动,以及在不同时间和领域内的变化,有助于捕捉科技公司与国家治理关系中的动态性。

表 1　　科技公司三种角色间的区别

		类别				
		原因	特征	表现形式	结果	研究议题
角色	治理主体	重构经济社会组织方式	主体性	数字空间中的基础性地位	科技公司在国家治理中的影响增强	塑造新型的权力形态结构
	治理工具	掌握先进社会治理技术	功能性	政府购买服务的承接主体	各级政府的技术治理水平显著提升	国家治理变革的科技支撑
	治理对象	蕴含复杂社会风险现象	风险性	新型国家治理问题的涌现	各级政府的治理体系发生深刻变革	治理挑战与治理改革方向

从治理主体的角度来看,科技公司自身重构了经济社会的运行方式,成为一种新的组织方式。它能够依托自身的资源自主发现社会问题并提供技术解决路径,凸显出其主体性特征,从而顺利地参与到数字时代的公共事务治理中。伴随着科技公司在数字空间中所处的基础性地位,以及国家与公众对它愈加深入的依赖程度,科技公司在国家治理中作为主体的影响力显著增强。现有研究议题多关注科技公司介入国家治理过程中,通过其垄断性的技术、数据等资源要素提升自身的权力地位,塑造出一种新的权力形态,改变既有的权力结构。

从治理工具的角度来看,在作为治理主体的科技公司实践中,科技公司因自身所掌握的先进社会治理技术而具有工具属性,其功能性特征能够有效地实现既定政策目标。通过政府购买公共科技服务参与国家治理,科技公司为政府提

供数字技术、智能化系统和数据分析等支持，显著提升了各级政府的技术治理水平，推动了国家治理的现代化和数字化转型。现有研究多从科技公司在国家治理变革中提供的技术支持切入，突出其作为治理工具的辅助性作用。

从治理对象的角度来看，国家之所以需要对科技公司进行治理，是因为科技公司所引发的广泛影响蕴含着复杂的社会风险现象，威胁到了社会的稳定和安全。科技公司所具备的风险性特征也使各种前所未有的社会问题层出不穷，这迫切地需要各级政府通过制定政策、法律法规和更新监管体系等治理变革来予以应对。现有研究已关注到这种迫切，集中分析科技公司或数字平台引发的治理问题，以及国家能够进行完善与改革的方向。

尽管我们提供了“治理主体—治理工具—治理对象”的三维分析框架，但这只是一种理想形式的讨论，并不意味着现实中这三种角色是互相分立、毫无关联的。实际上，在现实互动中，这三种角色之间存在着相互关联。更为复杂的是，它们之间可能涌现出各种组合形式。以科技公司为例，它有可能兼具这三种角色，而如何巧妙应对这些角色之间的潜在张力成为关键问题。如图 1 所示，解决之道或在于意识到这三种角色并非静态存在，它们在互动中是可以相互转化和调整的。

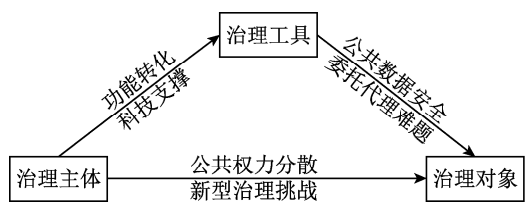


图 1　科技公司三种角色间的联系

（一）治理主体和治理对象的关系

首先，当科技公司作为治理主体时，其实质是在分享公共权力。当治理场域由实体空间转移至数字空间，由于拥有对技术的掌握能力，科技公司成为数字空间中社会生活的影响主体。国家作为公共事务治理的主体，可能面临技术不足以满足治理需求的挑战。为了实现国家治理现代化的目标，政府需依赖科技公

司的资源和专长来实现治理转型。同时,社会公众也越来越依赖在线平台进行社交、交易以及与政府部门互动。这使得公共权力被分散,政府在某些领域的决策能力被削弱,对科技公司产生过度依赖,从而影响政府的自主性和治理的独立性,所以需要对科技公司展开治理。

其次,科技公司的崛起给传统治理结构带来了前所未有的新型挑战和风险。这些公司凭借自身强大的技术创新能力,在短时间内积累了巨大的经济基础和社会影响力。通过数据控制、信息流通以及算法决策等手段,科技公司影响着公众的生活方式、价值观念乃至社会秩序。然而,这种新兴力量的迅猛发展也暴露了众多社会风险,如个人隐私泄露、数据安全问题、信息不对称以及算法偏见等,这些问题不仅威胁到公众的权益,也对社会治理体系构成了挑战。

因此,科技公司从治理主体转变成了需要被监管和治理的对象。这种转变要求政府和社会各界重新审视和构建适应数字时代的治理机制,不仅要加强对科技公司的法律法规监管,还需推动这些公司承担起社会责任,确保技术发展与社会公益的和谐共生。

(二)治理主体和治理工具的关系

当科技公司作为治理工具参与国家治理时,它在某种程度上扮演着治理主体的角色,只不过这种主体性处于相对附属的地位。将科技公司的先进技术和大数据分析等能力用于国家治理,使原本处于私域的、具有市场性的功能扩展到公共领域,在社会治理中发挥越来越关键的作用,协助政府更高效地履行职责。科技公司介入公共事务治理,能够帮助政府更有效地响应社会需求、预测和防范风险,能够切实提升公共服务的质量和效率。

在政府购买服务的背景下,科技公司成为部分服务的承接者和提供者,如技术支持、数据分析、智能系统等服务。这是因为它作为治理主体所具备的独特技术和功能,使它同时能够为国家治理现代化提供科技支撑,以确保政府的数字化和智能化目标得以实现。值得一提的是,科技公司在这一过程中并非仅仅被动地执行任务,它还通过其专业知识和技术能力积极参与并塑造国家治理的方向。这一互动关系的特殊性在于,科技公司与政府之间的合作不仅仅是一种商业关系,更是一种社会契约。科技公司不限于只赚取利润,更偏向于作为国家治理体

系的一部分。正因如此,科技公司从一个独立的治理主体变成国家进行治理时所采用的一种得力工具。

(三)治理工具和治理对象的关系

当科技公司充当治理工具的角色时,其所拥有的先进技术能够帮助政府提高治理效率,优化公共服务,推动国家治理转型。对这些资源的有效运用满足了国家面对虚拟空间的治理需求,因而科技公司得到了国家的广泛关注和认可。为了有效地提供数字技术和支持政府的决策制定,科技公司通常需要接触大量的公共数据,政府所掌握的这些公共数据将会暴露在公司层面,这本身带来了一系列潜在风险。公共数据包含大量敏感信息,数据泄露不仅会导致隐私侵犯,还可能危害国家安全,由此引发出对科技公司的监管和治理。如何确保数据在使用过程中的安全,成为政府在使用科技公司作为治理工具时必须面对的挑战。

再者,委托代理难题是另一个需要关注的问题。科技公司在追求技术进步和商业利益的过程中,所提供的服务和技术可能并不完全符合公共利益或治理伦理。如果没有有效的监管和控制机制,这些工具可能会在利益驱动下偏离原本的治理目标甚至被滥用,对社会造成负面影响。最重要的是,科技公司借助国家委托之便滥用公共数据和权力可能损害公众信任,并引发社会不满。因此,当科技公司充当治理工具时,虽然可以为国家提供重要的支持,但国家也应认真应对数据安全、操作和道德风险等问题。

由此,相比于既往研究关注科技公司参与到国家治理中,通过与国家或社会互动所引发出的新的治理问题,本文所提出的"治理主体—治理工具—治理对象"三维分析框架更注重科技公司所扮演的角色本身对经济、社会造成的影响及其所带来治理问题的情况。正是因为明确了三种不同的角色,我们得以深入探讨它们之间的差异与联系。这一研究视角为我们提供了重要的分析框架,有助于我们更好地理解每个研究议题的定位,也能够凸显已有研究中容易疏忽的问题领域,为未来的研究提供了借鉴。

科技公司与国家治理之间的关系,有着充分的研究空间。随着数字时代的发展,科技公司将发挥越来越重要的作用,为国家治理提供创新性解决方案。本文所提出的"治理主体—治理工具—治理对象"三维分析框架为政策制定者提

供了更全面的视角,通过分析不同维度,政策制定者可以更好地调整政策以更有效地管理和引导这一关系,并为研究者提供深入研究的方向,推动对这一重要议题更加深刻的理解。然而,关于科技公司参与社会治理的研究还有待进一步挖掘。比如,本文更偏重于对中国科技公司参与国家治理的考察,而比较不同国家和地区对科技公司在国家治理中角色的看法和实践,将有助于理解文化、法律和政治环境对科技公司治理作用的影响,并为国际合作和经验交流提供有益的经验;通过研究政府各部门之间数据碎片化的原因与改进策略,可能能够进一步解决数据协同共享与整合的问题,推进数据驱动的治理;除政府对平台的监管外,或许我们可以把目光聚焦于平台内部,关注平台自身如何通过自我约束和规范来适应国家和社会公众的期待,从而提升其在国家治理中的作用。

参考文献

艾云、李祥、向静林,2023,《数字时代金融诈骗组织的运作机制——基于非矿公司金融诈骗的网络民族志研究》,《社会学研究》第 5 期。

艾云、向静林,2021,《分布式网络、复杂性与金融风险治理——一个多层级政府行为的分析框架》,《社会学评论》第 1 期。

鲍静、范梓腾、贾开,2020,《数字政府治理形态研究:概念辨析与层次框架》,《电子政务》第 11 期。

北京大学课题组、黄璜,2020,《平台驱动的数字政府:能力、转型与现代化》,《电子政务》第 7 期。

财政部,2018,《政府购买服务管理办法(征求意见稿)》,https://www.sohu.com/a/238163302_99897000。

曹堂哲、魏玉梅,2019,《政府购买服务中的绩效付酬:一种公共治理的新工具》,《改革》第 3 期。

陈鹏,2019,《人工智能时代的政府治理:适应与转变》,《电子政务》第 3 期。

陈荣卓、肖丹丹,2015,《从网格化管理到网络化治理——城市社区网格化管理的实践、发展与走向》,《社会主义研究》第 4 期。

陈涛、董艳哲、马亮等,2016,《推进“互联网+政务服务”提升政府服务与社会治理能力》,《电子政务》第 8 期。

陈伟光、明元鹏,2021,《数字货币:从国家监管到全球治理》,《社会科学》第 9 期。

陈伟光、钟列炀、裴丹,2022,《平台经济治理:从国家监管向全球治理跨越的逻辑》,《天津社会科学》第 6 期。

樊鹏等,2020,《新技术革命与国家治理现代化》,北京:中国社会科学出版社。

冯乾、王海军,2017,《互联网金融不当行为风险及其规制政策研究——以市场诚信、公平竞争与消费者保护为核心》,《中央财经大学学报》第 2 期。

高海涛,2021,《我国企业如何参与互联网治理——基于百度、阿里和腾讯(BAT)的案例研究》,《新闻与传播研究》第 2 期。

高奇琦,2020,《智能革命与国家治理现代化初探》,《中国社会科学》第 7 期。

高奇琦,2023,《国家数字能力:数字革命中的国家治理能力建设》,《中国社会科学》第 1 期。

关婷、薛澜、赵静,2019,《技术赋能的治理创新:基于中国环境领域的实践案例》,《中国行政管理》第 4 期。

国家互联网信息办公室,2022,《数字中国发展报告(2021 年)》,http://www.cac.gov.cn/rootimages/uploadimg/1675765283208335/1675765283208335.pdf?eqid=cb0776d800041f8400000006643deb35&wd=&eqid=abaec189000359a9000000066575c6c3。

国务院,2015,《促进大数据发展行动纲要》,https://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm。

胡建绩、孔德洋,2000,《高新技术企业科研开发组织的模式选择》,《中国软科学》第 12 期。

华子岩,2023,《政府数据共享视域下首席数据官制度的确立及其风险防范》,《中国科技论坛》第 9 期。

黄波,2010,《私营企业主阶层的非制度性政治参与路径研究》,《经济体制改革》第 4 期。

黄冬娅、杜楠楠,2022,《平台企业政府事务部门专门化与政企关系发展——基于国家制度环境的分析》,《社会学研究》第 6 期。

黄其松,2022,《数字时代的国家理论》,《中国社会科学》第 10 期。

吉鹏、许开轶,2020,《大数据驱动下政府购买公共服务精准化:运行机理、现实困境与实现路径》,《当代世界与社会主义》第 3 期。

江小涓、黄颖轩,2021,《数字时代的市场秩序、市场监管与平台治理》,《经济研究》第 12 期。

姜红德,2020,《从“浙里办”看省级政务 APP 建设》,《中国信息化》第 8 期。

金兼斌,2019,《权力的游戏:资本、技术及其他》,《人民论坛·学术前沿》第 14 期。

金璇、赵玉现,2017,《共治视角下的网络谣言分析——微信平台谣言治理实践》,《新闻与写作》第 6 期。

蓝江,2022,《数据—流量、平台与数字生态——当代平台资本主义的政治经济学批判》,《国

外理论动态》第1期。

雷鑫洪,2021,《智慧税警协作下的涉税犯罪治理研究》,《税务与经济》第1期。

李洪佳,2020,《区块链技术助推政府购买公共服务的优势及建议》,《理论导刊》第11期。

李凌浩、吕鹏、刘金龙,2022,《中国县域中小企业数字化转型报告》,《中国经济报告》第6期。

李明,2015,《社交媒体视阈下的城市传播研究》,《中国出版》第13期。

吕鹏、房莉杰,2020,《寻找座头鲸:中国企业是如何进行社会创新的?》,北京:中国社会科学出版社。

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。

梅立润,2023a,《技术置换权力:人工智能时代的国家治理权力结构变化》,《武汉大学学报》(哲学社会科学版)第1期。

梅立润,2023b,《技术竞争与权力流动:人工智能时代国家治理权力的空间配置变化》,《哈尔滨工业大学学报》(社会科学版)第3期。

孟天广,2021,《政府数字化转型的要素、机制与路径——兼论"技术赋能"与"技术赋权"的双向驱动》,《治理研究》第1期。

孟天广、赵娟,2018,《大数据驱动的智能化社会治理:理论建构与治理体系》,《电子政务》第8期。

乔天宇、邱泽奇,2020,《复杂性研究与拓展社会学边界的机会》,《社会学研究》第2期。

邱泽奇、乔天宇,2021,《电商技术变革与农户共同发展》,《中国社会科学》第10期。

容志,2019,《大数据背景下公共服务需求精准识别机制创新》,《上海行政学院学报》第4期。

单勇,2019,《跨越"数字鸿沟":技术治理的非均衡性社会参与应对》,《中国特色社会主义研究》第5期。

单勇,2020,《犯罪之技术治理的价值权衡:以数据正义为视角》,《法制与社会发展》第5期。

单勇,2022,《数字平台与犯罪治理转型》,《社会学研究》第4期。

单勇,2023,《从传统犯罪学到数字犯罪学的代际更新》,《上海大学学报》(社会科学版)第4期。

史晋川、钱滔,2005,《制度创新与民营经济成长——以浙江台州为例的历史制度分析(HIA)》,《浙江学刊》第2期。

宋锴业,2020,《中国平台组织发展与政府组织转型——基于政务平台运作的分析》,《管理世界》第11期。

孙晋,2021,《数字平台的反垄断监管》,《中国社会科学》第5期。

孙晋、蔡倩梦,2023,《公平竞争原则下数字平台治理的规则补正》,《财经法学》第1期。

王锋,2022,《从治理到被治理:论基于数字平台型企业的社会治理》,《浙江学刊》第4期。

王伟域,2022,《数字文明时代税收治理现代化探析——基于区块链技术应用于税收治理的视角》,《理论月刊》第12期。

王晓燕,2006,《私营企业主非正式政治参与的途径与意义分析》,《南京师大学报》(社会科学版)第6期。

魏小雨,2019,《政府主体在互联网平台经济治理中的功能转型》,《电子政务》第3期。

吴晓波、陈颖,2014,《中小企业组织二元性对企业绩效的影响机制研究》,《浙江大学学报》(人文社会科学版)第5期。

向静林,2021,《互联网金融风险与政府治理机制转型》,《社会科学研究》第1期。

向静林、钟瑞雪,2024,《数字时代的地方金融风险与政府治理变革——"不完全发包"概念的引出》,《浙江学刊》第2期。

谢平、邹传伟,2012,《互联网金融模式研究》,《金融研究》第12期。

谢新洲、石林,2020,《基于互联网技术的网络内容治理发展逻辑探究》,《北京大学学报》(哲学社会科学版)第4期。

谢新洲、石林,2021,《国家治理现代化:互联网平台驱动下的新样态与关键问题》,《新闻与写作》第4期。

解正山,2020,《数据泄露损害问题研究》,《清华法学》第4期。

新华社,2013,《中共中央关于全面深化改革若干重大问题的决定》,https://www.gov.cn/zhengce/2013-11/15/content_5407874.htm。

熊鸿儒,2019,《我国数字经济发展中的平台垄断及其治理策略》,《改革》第7期。

徐汉明、张新平,2018,《网络社会治理的法治模式》,《中国社会科学》第2期。

徐家良、赵挺,2013,《政府购买公共服务的现实困境与路径创新:上海的实践》,《中国行政管理》第8期。

许晓龙,2015,《网络公共事件中的"泛政治化"现象探微》,《电子政务》第7期。

阳镇,2018,《平台型企业社会责任:边界、治理与评价》,《经济学家》第5期。

杨明,2022,《平台经济反垄断的二元分析框架》,《中外法学》第2期。

杨燕英、胡文辉、田开学,2019,《政府购买公共科技服务运行机制研究》,《中国政府采购》第7期。

杨志云,2019,《社会治安的政治定位与调控中的社会秩序——当代中国警务运行机理的解释框架》,《社会学研究》第2期。

叶娟丽、徐琴,2019,《去中心化与集中化:人工智能时代的权力悖论》,《上海大学学报》(社会

科学版)第 6 期。

郁建兴、樊靓,2022,《数字技术赋能社会治理及其限度——以杭州城市大脑为分析对象》,《经济社会体制比较》第 1 期。

占佳,2021,《互联网金融反垄断规制政策研究》,《新金融》第 9 期。

张丙宣、华逸婕,2019,《共享经济的监管:一个分析框架——以共享单车为例》,《浙江社会科学》第 5 期。

张茂元,2021,《技术红利共享——互联网平台发展的社会基础》,《社会学研究》第 5 期。

赵汀阳,2022,《假如元宇宙成为一个存在论事件》,《江海学刊》第 1 期。

周飞舟,2019,《政府行为与中国社会发展——社会学的研究发现及范式演变》,《中国社会科学》第 3 期。

周辉宇,2017,《基于大数据规则挖掘的交通拥堵治理研究》,《统计与信息论坛》第 5 期。

周俊,2014,《政府如何选择购买方式和购买对象? ——购买社会组织服务中的政府选择研究》,《中共浙江省委党校学报》第 2 期。

周黎安,2017,《转型中的地方政府:官员激励与治理》,上海:格致出版社。

周文、韩文龙,2021,《平台经济发展再审视:垄断与数字税新挑战》,《中国社会科学》第 3 期。

周雪光,2008,《基层政府间的"共谋现象"——一个政府行为的制度逻辑》,《社会学研究》第 6 期。

朱光磊、杨立武,2004,《中国私营企业主政治参与的形式、意义和限度》,《南开学报》第 5 期。

朱火弟、蒲勇健,2003,《政府绩效评估研究》,《改革》第 6 期。

Andrejevic, M. 2014, "The Big Data Divide." *International Journal of Communication* 8.

Ansell, C. & A. Gash 2018, "Collaborative Platforms as A Governance Strategy." *Journal of Public Administration Research and Theory* 28(1).

Ansell, C. & S. Miura 2019, "Can the Power of Platforms Be Harnessed for Governance?" *Public Administration* 98(1).

Chesbrough, H. 2003, "The Era of Open Innovation." *MIT Sloan Management Review* 44(3).

Danah, B. & K. Crawford 2012, "Critical Questions for Big Data: Provocations for A Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15.

Huang, D. & M. Chen 2020, "Business Lobbying Within the Party State: Embedding Lobbying and Political Co-Optation in China." *China Journal* 83(1).

Janssen, M. & E. Estevez 2013, "Lean Government and Platform-Based Governance: Doing More with Less." *Government Information Quarterly* 30(4).

Jungherr, A. 2016, "Twitter Use in Election Campaigns: A Systematic Literature Review." *Journal of Information Technology & Politics* 13(1).

Tushman, M. & C. O'Reilly 1996, "Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change." *California Management Review* 38.

新闻大数据视角下测度区域一体化水平的新路径

——以长三角地区为例

李　军　董方杰*

摘要：本文是利用新闻大数据对区域一体化水平进行测度的首次尝试。本文采集并处理了长三角地区41座城市的新闻媒体在2016—2022年发布的所有新闻报道，利用大数据文本分析和社会网络分析技术测度长三角一体化水平，并以长三角一体化指数呈现历时态（时间）与共时态（空间）下一体化的趋势与差异，探索了一条借助新闻大数据测度区域一体化水平的新路径。

关键词：新闻大数据　社会网络分析　长三角一体化　长三角城市群

一、引言

城市化是经济发展的关键动力（Bairoch，1991；Bertinelli & Black，2004），对现代社会的基本架构产生了深远影响。经济地理学揭示了城市作为各类生产要素聚集中心的重要性（Storper，2013）。尤其是自20世纪中叶以来，随着经济全球化的快速推进，区域一体化的进程也在不断加速。在民族国家内部，出现了一种新型区域联合体，被称为“城市群”（megalopolis），它展示了经济和城市发展从点到面的跨越，也标志着城市化进入了更为发达的空间形态（Fang & Yu，2017），从相互竞争转变为协同发展（Fang & Yu，2020）。

全球范围内有五个被广泛承认的世界级城市群，分别是美国东北部大西洋沿岸城市群、北美五大湖城市群、日本太平洋沿岸城市群、英伦城市群以及欧洲

* 李军，南京大学社会学院。董方杰，南京大学社会学院。

西北部城市群(Gottmann, 1976)。随着中国改革开放后城市化的快速推进,城市群逐渐成为中国新一轮城镇化的重要推动力(Zhao, Zhang & Zhao, 2021)。目前,长三角城市群已成为六个世界级城市群之一(方创琳、姚士谋、刘盛和等,2011;胡雅龙,2010),充分展示了长三角城市群在经济和战略上的重要地位。

相较于其他世界级城市群,长三角城市群展现出明显的国家规划特色。2010 年 5 月,中国批准实施《长江三角洲地区区域规划》,正式确定了长三角地区的范围为江浙沪,并明确了其发展战略定位:成为亚太地区重要的国际门户、全球重要的现代服务业和先进制造业中心,以及具有较强国际竞争力的世界级城市群。该规划主要集中在区域规划方面,核心内容是划定各城市的发展重点。其目的在于避免这些城市之间的恶性竞争,明确它们各自的产业定位,以解决区域城市发展过程中可能出现的同质化问题。然而,该规划并未充分体现对区域一体化和协同发展的思考。

直至 2018 年 6 月,国家主席习近平在首届中国国际进口博览会上宣布支持长江三角洲地区的一体化发展,并将其提升为国家战略。紧随其后,2019 年 12 月 1 日,中共中央、国务院印发《长江三角洲区域一体化发展规划纲要》作为指导长三角地区当前及今后一个时期一体化发展的纲领性文件,这是中央层面首次将安徽省整体纳入长三角城市群的范畴。文件指出,长三角地区是我国经济发展最活跃、开放程度最高、创新能力最强的区域之一,在国家现代化建设大局和全方位开放格局中具有举足轻重战略地位。推动长三角一体化发展,增强长三角地区的创新能力和竞争能力,提高经济集聚度、区域连接性和政策协同效率,对于引领全国高质量发展、建设现代化经济体系具有重大意义。在产业集聚效应和政策规划引领的双重推动下,长三角城市群的一体化进程进入了一个新的发展阶段。

与此同时,学界对于测度区域一体化水平的讨论正在蓬勃发展,主要有两条研究路径。其中一条路径遵循传统的经济学研究范式,另一条则源自网络社会的崛起和“流动空间”(space of flow)(Castells, 2000)概念的提出。传统测度区域一体化水平的路径主要使用现有的经济统计数据,尽管在某种程度上可以反映区域一体化的程度,但这些统计数据存在一个重要问题,即

缺乏方向性尺度，无法真实地测量区域内各城市因各类互动而形成的网络关系。不过，近年来大数据研究的兴起为弥补上述研究的缺陷提供了可能性。

二、文献回顾

本文的主要目的是利用新闻大数据的方法，为测度区域一体化水平提供一条新的研究路径。在过去关于测度区域一体化水平的研究中，可以明显看出两条清晰的研究路径，这也反映了区域研究一体化的范式转型。

第一条研究路径采用传统的经济学研究方法。该路径利用所在区域的各类社会经济指标，主要考察人口、GDP、财政支出等经济活动在空间距离上的集聚。这类研究路径通常将属性数据（data of city attributes）（李涛、周锐、苏海龙等，2015；Beaverstock，Smith & Taylor，2000；Taylor，1999）按研究重点划分为不同的指标维度，然后借助各类经济学模型来测度区域经济一体化水平。例如，引力模型（gravity model）广泛应用于城市之间联系强度的测算（李涛、周锐、苏海龙等，2015；许静涵、林礼娜、武义天等，2021；Chen，2009；Chen & Huang，2018；Fang，Sheng & Li，2018；Han，Cao & Liu，2018；Zhao，Zhang & Zhao，2021），也可用于测算边界效应，衡量地区边界导致市场分割的程度（李晓、王小彬，2017；王晔倩、林理升，2006；肖海平、谷人旭，2011；行伟波、李善同，2009；赵永亮、徐勇、苏桂富，2008；赵永亮、才国伟，2009；Broker，1984；Engel & Rogers，1994；McCallum，1995；Poncet，2003）；拟合分析模型（邓淑芬、朱佳翔、钟昌宝，2018；刘干、陆叶，2018；Ibrahim，2008；Kim，2001；König & Ohr，2013；Wenjie，2022）通常使用聚类方法对各类指标进行层次化，以得出城市群一体化发展水平的差异。具体到长三角一体化水平的相关研究中，孙红梅等人采用熵值法构建了 2010—2017 年长三角一体化协同评价指标体系，并通过全局自相关与局部空间自相关散点分析发现长三角一体化协同具有较强的空间正相关性（孙红梅、姚书淇，2020）；刘志彪等人提出了一种综合指标体系，涵盖了十个维度，即空间、市场、产业、创新、交通、信息、贸易、公共服务、生态环境和制度，以此来测量长三角一体化程度（刘志彪、孔令池，2019）；李世奇等人采用基于专家咨

询的主成分分析评价方法构建了长三角一体化指标体系，该方法侧重于市场统一性、要素同质性、发展协同性和制度一致性，为理解长三角一体化程度提供了新的视角（李世奇、朱平芳，2017）。

第二条研究路径源自卡斯特尔（M. Castells）在《网络社会的崛起》一书中提出的“流动空间”概念。他认为，社会由资本、信息、技术等“流”组成，城市网络位于流动空间中。换言之，城市之间的网络关系而非等级结构塑造了世界和区域城市网络，使得流动空间成为信息社会的主导性空间组织形式（Castells，2000）。这一观点直接推动了城市研究从场所空间（space of place）转向流动空间（space of flow），城市网络的研究范式得到了众多学者的关注（Beaverstock，Smith & Taylor，2000；Taylor，1999）。为了弥补属性数据无法动态展现城市网络关系的缺陷，学者们开始使用关系型数据（relational data）。在测度区域一体化水平与城市网络关系方面，国内外学者运用交通流（physical transportation）（Derudder & Witlox，2005；Mahutga，Ma & Smith et al.，2010；Smith & Timberlake，2002；Van Nuffel，Derudder & Witlox，2010）、贸易流（Afesorgbor，2017；Lee，Owen & Mensbrugghe，2009；Rossi，Beaverstock & Taylor，2007；Wilson & Otsuki，2007）和电信流（Graham & Marvin，1995；Hall & Pain，2006；Rutherford，Gillespie & Richardson，2004）等流量数据指标来进行。

在长三角城市群研究中，不同学者采用了不同方法来探索该地区的特点。朱查松等人使用2001年、2009年长三角企业分支机构数据，通过分析企业联系网络来研究长三角城市网络的区域空间组织特征（朱查松、王德、罗震东，2014）。唐子来等人基于企业区位数据，采用总部分支法，分析了京津冀、长三角和珠三角地区的城市关联网络。他们发现，北京、上海、广州和深圳不仅是中国的核心城市，还分别是三个地区的“门户城市”，承担着连接全球经济网络和辐射区域腹地的双重作用（唐子来、李涛，2014）。李涛等人以关系型大数据为基础，选择了企业分支数据、信息流数据和交通流数据，运用关联网络分析和多变量分析方法，对长三角地区30个城市的区域经济一体化指数进行了测量（李涛、周锐、苏海龙等，2015）。这种方法不仅考虑了企业间的联系，还综合考虑了信息和交通流动，为评估长三角地区的经济一体化程度提供了多维度视角。此外，一些学者还通过探索城市网络结构（唐子来、赵渺希，2010；赵渺希、唐子来，

2010）、创新协同网络（白俊红、蒋伏心，2015；徐宜青、曾刚、王秋玉，2018；张贵、温科，2017）等方式，对长三角城市群进行了研究。

近年来，大数据研究的兴起为我们从流动空间的角度更细致、翔实地理解区域一体化进程提供了新的视角。大数据的优势在于充分利用互联网中的大量数据，以近乎全样本的方式对社会事实进行全面观察、计算和分析，为研究大空间的宏观或微观经济现象提供了全新的路径。虽然曾有学者从事过相关研究，如熊丽芳等人采用创新方法构建了中国东部、中部和西部三大经济区的城市网络。他们利用百度指数中的用户关注度数据，反映了经济区内各城市之间的互联互通程度。通过模拟计算百度信息流，可以描绘出中国三大经济区内城市之间的网络关系（熊丽芳、甄峰、王波等，2013；熊丽芳、甄峰、席广亮等，2014）。甄峰等人通过分析微博用户数据，探究了微博社会空间视角下的城市网络结构，发现中国城市网络在微博平台上呈现明显的等级关系和层级区分，表明城市的网络连接度与城市等级存在一定的相关性，揭示了城市在数字空间中的社会互动模式，为理解城市间的互联互通和影响力分布提供了新的路径（甄峰、王波、陈映雪，2012）。然而，这种基于个体行为的大数据虽然能够从某个侧面“复刻”区域一体化的程度，但无法完整呈现区域一体化的真实实践，而新闻大数据则弥补了上述研究的不足。

三、数据来源与研究方法

（一）新闻大数据的数据来源

本文所采用的数据来源为锐研中文新闻数据库。作为国内学术领域知名的中文新闻数据库之一，该数据库目前覆盖了全网各类电子报纸、政府官方新闻、大型新闻门户等超过 1300 家互联网新闻媒体来源的全文索引数据。该数据库还提供了新闻内容主题分类，并能提取各类结构化变量，如时间、地点、人物、机构、主题、情感和影响力。借助该数据库，我们可以进一步搜集指定媒体范围和时间范围的全文数据。

长三角城市群涉及三省一市，共有 40 个地级市和 1 个直辖市。本文选取

的媒体主要为各地级市的日报，如《南京日报》《杭州日报》《合肥日报》；个别城市的日报不以城市名称命名，如六安《皖西日报》、宿州《拂晓报》。此外，上海市选取了一份日报《文汇报》和一份晚报《新民晚报》，江苏省、浙江省和安徽省也各选取了一份省级日报和省级晚报。总计涉及 48 家报纸从 2016 年 1 月 1 日至 2022 年 12 月 31 日的所有新闻报道数据，共计超过 505 万篇（表 1）。

表 1　　2016—2022 年长三角城市群新闻报道总数

年份	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	总计
新闻报道总数	616 565	633 153	748 157	778 321	860 171	719 992	703 459	5 059 818

我们从 505 万余篇新闻报道中进一步筛选出涉及超过 2 座城市的新闻文本，最终得到了 669 346 篇共现新闻报道作为研究的数据总集（表 2）。

表 2　　2016—2022 年长三角城市群共现新闻报道总数

年份	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	总计
共现新闻报道总数	89 582	94 540	107 156	113 110	107 248	85 573	72 137	669 346

（二）长三角一体化指数的构建策略

区域一体化发展首先在于加强区域内不同城市之间的连接，这种“连接”可以体现在多个方面，包括交通互联、文化交流、科创协同、招商引资、环保协作、医疗卫生和教育资源共享等。新闻是对新近发生、正在发生或仍能产生社会效应的事实进行的报道，城市之间的合作交流事件通常会被新闻所记录。当一篇新闻报道同时提及 2 个城市时，这两个城市往往存在某种“连接”。因此，我们可以通过文本分析技术统计一段时间内同时提及 2 个城市的文本次数，从而近似地测量这两个城市之间的绝对连接强度。表 3 展示了 2016 年上海与长三角各城市在新闻大数据集中的共现次数。

表 3 2016 年上海与长三角各城市在新闻大数据集的共现次数

城市	共现次数	城市	共现次数	城市	共现次数	城市	共现次数	城市	共现次数
南京市	7789	盐城市	895	嘉兴市	1714	合肥市	3327	黄山市	762
无锡市	2445	扬州市	1028	湖州市	1429	芜湖市	1265	滁州市	792
徐州市	1808	镇江市	985	绍兴市	1709	蚌埠市	499	阜阳市	488
常州市	1822	泰州市	1191	金华市	1236	淮南市	672	宿州市	275
苏州市	4469	宿迁市	469	衢州市	911	马鞍山市	562	六安市	462
南通市	1891	杭州市	9380	舟山市	815	淮北市	240	宣城市	592
连云港市	613	宁波市	3495	台州市	890	铜陵市	404	亳州市	250
淮安市	429	温州市	1728	丽水市	815	安庆市	742	池州市	389

如果一篇报道同时提及超过 2 个城市，那么每个城市两两之间都会被计算一次。考虑到新闻报道的总数量并不是恒定的（尽管报纸的版面总数大致相同，但文章数量通常会有变化），因此我们可以将某一年的绝对连接强度除以该年提及多个城市的报道总数（共现新闻报道总数），以得到相对连接强度。

然而，需要注意的是，并非每次新闻报道提及多个城市都意味着这些城市之间存在较强或明显的“连接”关系。例如，一篇报道可能只是介绍了一位来自江苏泰州的上海人。这类新闻数量相对较少，而且在一定程度上代表了两个城市之间的弱联系。对于海量的文本大数据而言，这种误差对最终的统计结果影响并不大。因此，使用上述统计方法来近似地测量不同城市之间的连接强度是可行的。

在特定区域内，我们可以将一个城市与该区域内其他城市的所有相对连接强度求和，以衡量该城市与整个区域的一体化程度；将所有城市与整个区域的融合程度再次汇总，就可以得到整个区域的一体化指数。

在一个具有 N 个城市的区域内，如果用 S_{ij} 来表示同时报道城市 i 和城市 j 的相对连接强度，用 CI_i 表示城市 i 与整个区域的一体化程度，用 RI 表示区域的整体一体化程度，则：

$$CI_i = \sum_{j=1}^{N} S_{ij}(j \neq i)$$

$$RI = 1/2 \sum_{i=1}^{N} CI_i$$

区域一体化指数 *RI* 的可解释意义在于：对于每一篇涉及多个城市的新闻报道，其平均包含的“边”（两两城市对）的数量。如果所有的新闻报道仅包含 2 个不同城市，则 *RI*=1.0。正是因为一篇新闻报道可能涉及 2 个以上的城市，“边”的数量会呈现非线性放大（3 个城市共有 3 条边，4 个城市共有 6 条边，5 个城市共有 10 条边，依此类推），因此 *RI*>1.0。如果单篇新闻报道普遍涉及越来越多的城市，则一定程度上说明城市之间的连接得到了加强，区域融合也更加深入。虽然单篇新闻平均报道的城市数量也可以用来作为区域融合度的测量指标，但“边”的平均数量更能反映其变化幅度，从而该指标具有更好的可表示性。

四、新闻大数据下长三角一体化进程的时空演变

（一）长三角一体化程度与共现新闻报道总数

为了研究长三角一体化程度的历时态变化情况，我们选择了 2016—2022 年连续 7 年的新闻大数据进行研究。以年为单位，首先计算了每年涉及多个城市的新闻报道占当年所有新闻报道的比例，以及长三角城市群的区域一体化指数。如图 1 所示，2016—2019 年的多城共现报道比例在波动中增大，表明长三角城市间互动的新闻比例在增加。这一现象在一定程度上验证了区域融合的持续发展，与长三角一体化指数的波动上升趋势相一致。2020 年后，多城共现报道比例有所下降，表明区域内城市间的合作和联动在新闻报道中的关注度有所降低。这一趋势在长三角一体化指数中也得到了验证。不过，令人欣喜的是，2022 年长三角一体化指数再次上升，表明区域整体融合再次展开。

（二）长三角各城市与区域一体化程度

长三角一体化指数的核心计算逻辑在于长三角 41 个城市之间的共现互动。

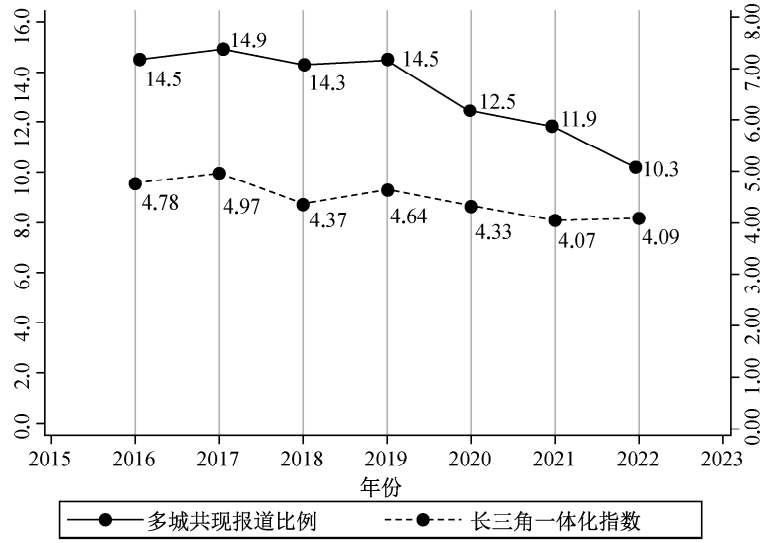


图 1　多城共现报道比例与长三角一体化指数

在前文中，我们定义了 CI_i 表示城市 i 与整个区域的一体化程度，接下来呈现的长三角各城市一体化指数，以 2016 年上海的一体化指数为基准（设为 100），进而标准化了 2016—2022 年各城市的指数值。我们采用空间分布和时间序列的双重维度来展现这一指数，其中表 4 根据历年指数平均值划分了城市的层级。

从空间分布来看，上海的一体化指数显著高于其他城市，凸显了其在长三角一体化进程中的中心地位，可视作区域融合的引擎。此外，南京和杭州的指数紧随其后，成为长三角融合的两大支点，而合肥、苏州、宁波等城市则作为次中心地带，在整个一体化进程中扮演着重要角色。第四层级包括无锡、徐州等 7 座城市，第五层级包括扬州、安庆等 17 座城市，第六层级包括衢州等 11 座城市。可以看到，长三角地区的一体化逐渐形成了一个“中心—两翼—多次中心”的层级格局。在各省内部，江苏省呈现的地理分布趋势为各城市的一体化指数自南至北、自东至西逐渐降低，浙江省各城市一体化指数也呈现出自北向南、自东向西逐步降低的趋势，江浙两省一体化指数的分布差异展现的是上海的辐射带动作用。安徽则较为特殊，合肥—南京的夹心地带一体化指数显著高于其他地区，体现的是南京在安徽诸城中独特的城市影响力。从时间序列角度来看，中心地带

的城市如上海、南京、杭州、苏州、宁波等，其一体化指数稳中有进；边缘地带城市的指数波动较大。

表 4　　长三角各城市一体化指数一览表

长三角一体化指数等级	城市（一体化指数）
第一层级（100 以上）	上海市（107.8）
第二层级（70—100）	南京市（92.3）、杭州市（83.7）
第三层级（40—70）	合肥市（60.9）、苏州市（57.8）、宁波市（46.0）
第四层级（30—40）	无锡市（36.9）、徐州市（34.6）、芜湖市（32.7）、南通市（31.3）、温州市（30.8）、嘉兴市（30.4）、常州市（30.1）
第五层级（20—30）	扬州市（29.5）、安庆市（27.8）、绍兴市（27.7）、金华市（26.8）、湖州市（26.5）、滁州市（26.1）、马鞍山市（25.8）、盐城市（25.0）、镇江市（24.8）、阜阳市（24.5）、宣城市（23.4）、台州市（23.3）、泰州市（22.2）、六安市（22.1）、蚌埠市（21.7）、黄山市（21.3）、淮南市（20.6）
第六层级（0—20）	衢州市（19.7）、连云港市（19.4）、淮北市（19.4）、丽水市（19.0）、铜陵市（18.8）、宿迁市（18.6）、舟山市（17.9）、池州市（17.7）、宿州市（17.6）、淮安市（17.5）、亳州市（16.8）

（三）长三角一体化进程中的城市社会网络分析

在前文中，我们定义了 S_{ij} 来表示同时报道城市 i 和城市 j 的相对连接强度，因此通过 S_{ij} 建立 41 * 41 的网络共现矩阵来呈现长三角各城市之间的网络互动。[①] 我们将从网络密度、中心度、凝聚子群三个层面展开分析。

首先，从网络密度分析来看。网络密度衡量了网络中节点之间联系的密集程度。在数学上，它被定义为实际连接数与最大连接数之间的比率。在社会网络的实际运用中，网络密度越大，网络中节点之间的联系越紧密，网络对节点的影响可能性越大。一般来说，在真实的世界中，网络密度总是低于 0.5（Mayhew & Levinger, 1976）。在表 5 中，我们运用 Ucinet 对长三角各城市之间的共现互动网络进行了测算，发现长三角城市网络的密度一直居于高

① 期刊版面有限，无法全面呈现图片信息，有意者请向作者索要。

位,意味着其一体化程度很高。虽然在新冠疫情影响下有所波动,但依然稳步上升,展现了长三角各城市之间的融合韧性,一体化进程始终在稳步推进。

表 5　　2016—2022 年长三角各城市的共现网络密度

年份	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年
数值	0. 3439	0. 3610	0. 3561	0. 3427	0. 3341	0. 3256	0. 3841

其次,我们从四大中心度的角度去考察长三角城市网络中的各个节点(城市)。由于涉及城市众多,并且中心度的分析主要用于评估节点的重要性,因此我们在图 2 中列举了 2016—2022 年长三角城市中四大中心度位列前五的城市,事实上这也是长三角一体化指数最高的几个城市,验证了我们构建长三角一体化指数的科学性。

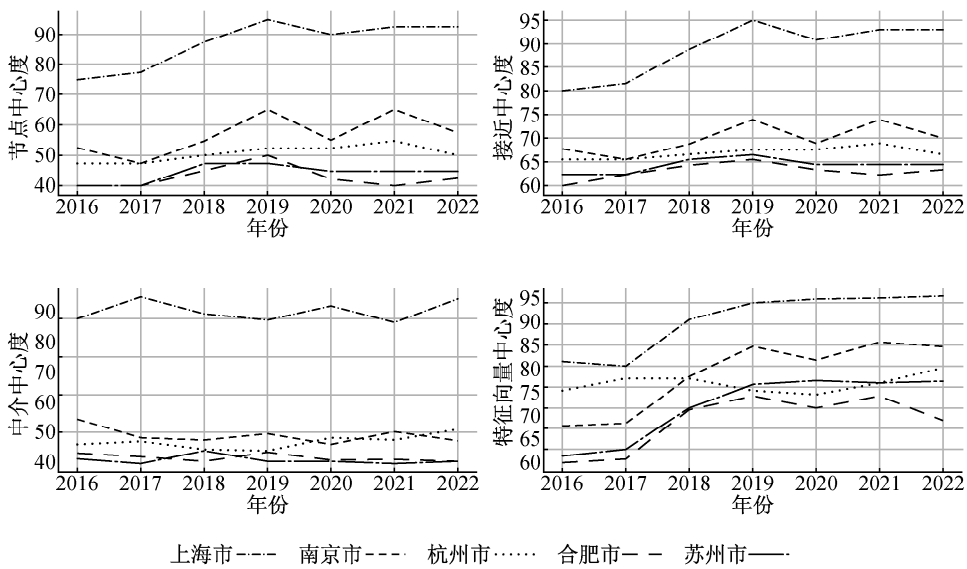


图 2　2016—2022 年长三角五大城市的中心度演变趋势

在社会网络分析中,通常使用节点中心度(degree centrality)或接近中心度(closeness centrality)来评估节点的重要性。节点中心度衡量了节点直接相连的数量,节点中心度越高,节点在网络中的重要性就越大,即位于网络的中心位置。然而,节点中心度忽略了节点之间的距离因素,因此引入接近中心度来弥补这个

局限。接近中心度通过计算节点到其他所有节点的最短距离来反映节点靠近中心的程度。接近中心度越高,节点在网络中的位置越靠近中心,其重要性也越显著。无论是节点中心度还是接近中心度,2016—2022 年期间,上海都是长三角地区最重要的网络节点,其次是南京、杭州、苏州和合肥。需要特别注意的是,2018—2019 年合肥的节点中心度和接近中心度有了显著提升,逼近甚至超过苏州。这说明自 2018 年以来,长三角一体化进程首次全面纳入安徽,以合肥为代表的安徽省域全面加强了与其他地区的联系与互动。从各城市的数值趋势来看,五大城市的节点中心度和接近中心度都呈增加趋势。这说明在长三角一体化大背景下,中心节点的功能性和重要性进一步增强。

此外,我们在评估某一特定节点的依赖性时,常使用中介中心度(betweenness centrality)作为评估标准,这类节点在实际生活中通常被称为“结构洞”。中介中心度的核心计算逻辑在于衡量节点在多大程度上扮演“中介”的角色。当一个节点经常出现在其他节点之间的最短路径上时,它的中介中心度就较高。这种中心度反映了节点在网络中的控制和影响力。中介中心度越高,节点在众多路径中越扮演必经之地,具有更强的传递或控制其他节点向外连接的能力。在长三角城市网络中,上海是最为重要的中介枢纽。南京和杭州之间存在着激烈的次中心竞争:2020 年之前,南京稳压杭州;2020 年之后,杭州的中介作用逼近甚至超过了南京。另外,在苏州和合肥之间的比较中,大部分年份中合肥的中介中心度都高于苏州。这是因为安徽的大部分城市需要依靠合肥作为枢纽来参与整个长三角的网络,而苏州因为与上海紧密相连,再加上南京作为重要的中介枢纽存在,导致苏州的中介能力相对较弱。同时我们还注意到,2020 年后合肥的中介中心度有所下降,这表明长三角一体化战略推动了安徽诸城市与长三角各城市之间的直接联系与互动。合肥中介中心度的下降,间接说明了长三角一体化程度的提高。

相较于前述三种中心度考虑数量和距离的因素,特征向量中心度(eigenvector centrality)更加注重节点连接的质量,而不仅仅考虑数量。特征向量中心度高的节点是那些连接了许多其他重要节点的节点,其重要性是通过相邻节点的重要性来评估的。从特征向量中心度的角度来看,上海依然是最为重要的城市,其次是南京、杭州、苏州和合肥。从数值的演变趋势来看,这五个主要城市的中

心度值大幅增加。这表明在长三角一体化进程中,各节点之间的联系质量进一步加强。

如果说我们通过四大中心度考察了重要中心节点的变化来侧面刻画长三角一体化的进程,那么进一步,我们用凝聚子群来完成“由点到面”的城市网络聚类。我们通过 Ucinet(最大分割深度为 2,集中标准为 0.2)对长三角城市网络进行了计算,最终将其划分为 4 个子群,分别为子群一(上海+江苏全域)、子群二(浙江全域)、子群三(合肥、芜湖、六安、宣城、马鞍山、池州、铜陵、安庆、滁州)、子群四(黄山、阜阳、宿州、蚌埠、淮南、亳州、淮北)。其中,在时空演变上,较大的变化在于 2016—2017 年,上海更多与浙江全域组成子群,但从 2018 年开始归类于江苏全域;相较于浙江、江苏较为稳定的子群结构,安徽全域则是一分为二,且这两个子群结构城市变化较大。①

五、总结与讨论

大数据研究方法的兴起,为以跨学科方式进行学术研究提供了可能(陈云松、吴青熹、黄超,2015; McFarland, Lewis & Goldberg, 2016; Ruppert, 2013)。这也导致了许多分支方向的出现,无论是计算社会学、计算经济学还是计算传播学,甚至更广义的“计算社会科学”,尽管它们在理论架构、计算方法和研究路径上存在差异,但核心思想都是基于大数据理念,挖掘传统定量研究无法触及的大量非结构化电子印迹数据(迈尔-舍恩伯格、库克耶,2013),以实现对社会现象和规律更精确的描述和解释(周涛、高馨、罗家德,2022)。本文正是基于这一核心思想和方法,收集并处理了长三角地区三省一市 41 座城市 2016—2022 年新闻媒体发布的所有报道。通过一系列数据处理方法,将百万级的非结构化数据转化为可进行有效分析的结构化数据。最终,我们成功构建了拥有 66 万余篇新闻文本的长三角新闻数据库,以测度长三角一体化水平。通过长三角一体化指数,我们呈现了长三角一体化在历时态(时间)和共时态(空间)下的趋势和差异。

① 期刊版面有限,无法全面呈现图片信息,有意者请向作者索要。

从方法层面来看，本文探索了一种新的途径，利用新闻大数据来测度区域一体化水平，以展示城市之间的实践和互动，并将其指数化、结构化和网络化。这也是计算传播学、经济地理学和社会网络分析的三重交叉尝试。从现实意义上讲，我们从时间和空间两个维度展示了长三角一体化的进程，为推动长三角一体化提供了重要的依据。研究发现，在整体一体化融合的趋势上，2017—2019 年长三角地区的一体化水平稳步递增，融合程度持续提高；然而，由于新冠疫情的影响，2020 年一体化程度有所减缓；但在 2022 年，这种下降趋势得到逆转，一体化指数再次上升，区域整体融合再次展开。这一点在后续的社会网络分析中得到了进一步验证，印证了长三角一体化指数构建的科学性。与此同时，本文对长三角 41 座城市进行了一体化水平的指数化操作。依据各城市的一体化指数，我们将长三角各城市划分为六大层级，可以看到长三角地区的一体化逐渐形成了一个“中心—两翼—多次中心”的层级格局。这样的层级格局体现为两个方面：首先是核心城市如上海、南京、杭州、苏州和合肥，在长三角网络节点中愈发重要的定位，这一点我们通过四大中间度进行了说明；其次在城市网络的内部子群中，核心城市主导了长三角内部子群的划分，由于核心城市的重要性愈发突出，长三角一体化的进程逐渐从“点—点”互动进程转向了“面—面”互动的子群融合。从政策后续来说，习近平总书记在深入推进长三角一体化发展座谈会上指出：“长三角一体化发展有许多深层次问题有待进一步破解，发展质量效率和辐射带动作用仍需提升，重点领域、重点区域一体化尚需努力，产业链供应链分工协作水平有待提升，建立全国统一大市场的龙头带动作用有待进一步发挥，改革开放还需进一步向纵深拓展，超大特大城市治理和发展还有不少短板。”本文认为，如果要推动长三角一体化往深层次发展，必须要利用长三角“中心—两翼—多次中心”的城市层级格局，以龙头带动高质量发展，以两翼衔接重点区域，以多次中心提升一体化水平，完成城市连接实现从“点—点”到“面—面”的过渡。

最后，本文在国内属于首次利用新闻大数据对区域一体化水平进行测度，得到的各城市一体化指数具有很好的后续研究价值。虽然在本文中，我们只对一体化指数进行了描述性的研究，但在接下来的长三角一体化讨论中，我们将利用各城市的一体化指数和各类政治、经济、社会、文化和环境等维度变量构建一个

多年面板数据。通过运用计量经济学模型，我们将深入探讨影响长三角一体化水平的因素，以提供更为实际的研究结论和政策建议。

参考文献

白俊红、蒋伏心，2015，《协同创新、空间关联与区域创新绩效》，《经济研究》第 7 期。

陈云松、吴青熹、黄超，2015，《大数据何以重构社会科学》，《新疆师范大学学报》（哲学社会科学版）第 3 期。

邓淑芬、朱佳翔、钟昌宝，2018，《长三角城市群物流产业空间关联格局演变——基于聚类、探索性空间数据和 σ 收敛性的分析》，《商业经济研究》第 20 期。

方创琳、姚士谋、刘盛和等，2011，《2010 中国城市群发展报告》，北京：科学出版社。

胡雅龙，2010，《世界第六大城市群：长江三角洲城市群崛起之路》，上海：上海社会科学院出版社。

李世奇、朱平芳，2017，《长三角一体化评价的指标探索及其新发现》，《南京社会科学》第 7 期。

李涛、周锐、苏海龙等，2015，《长三角区域经济一体化水平的测度：以关系型大数据为基础》，《新常态：传承与变革——2015 中国城市规划年会论文集》。

李晓、王小彬，2017，《粤港澳经济一体化：基于边界效应视角的分析》，《湖北社会科学》第 11 期。

刘干、陆叶，2018，《我国区域经济发展水平的聚类分析》，《生产力研究》第 1 期。

刘志彪、孔令池，2019，《长三角区域一体化发展特征、问题及基本策略》，《安徽大学学报》（哲学社会科学版）第 3 期。

迈尔-舍恩伯格、库克耶，2013，《大数据时代：生活、工作与思维的大变革》，盛杨燕、周涛译，杭州：浙江人民出版社。

孙红梅、姚书淇，2020，《长三角城市群一体化协同发展的空间实证研究——以财政水平差异分析为支撑》，《城市发展研究》第 8 期。

唐子来、李涛，2014，《京津冀、长三角和珠三角地区的城市体系比较研究——基于企业关联网络的分析方法》，《上海城市规划》第 6 期。

唐子来、赵渺希，2010，《经济全球化视角下长三角区域的城市体系演化：关联网络和价值区段的分析方法》，《城市规划学刊》第 1 期。

王晔倩、林理升，2006，《引力模型与边界效应分析——以长三角和珠三角服务贸易为例》，《上海经济研究》第 8 期。

肖海平、谷人旭,2011,《沪苏浙边界区县域经济合作与协调发展中的边界效应实证研究》,《经济研究导刊》第15期。

行伟波、李善同,2009,《本地偏好、边界效应与市场一体化——基于中国地区间增值税流动数据的实证研究》,《经济学(季刊)》第4期。

熊丽芳、甄峰、王波等,2013,《基于百度指数的长三角核心区城市网络特征研究》,《经济地理》第7期。

熊丽芳、甄峰、席广亮等,2014,《我国三大经济区城市网络变化特征——基于百度信息流的实证研究》,《热带地理》第1期。

徐宜青、曾刚、王秋玉,2018,《长三角城市群协同创新网络格局发展演变及优化策略》,《经济地理》第11期。

许静涵、林礼娜、武义天等,2021,《基于引力模型和聚类分析的城市群经济结构研究——以长三角23市为例》,《浙江金融》第2期。

张贵、温科,2017,《协同创新、区域一体化与创新绩效——对中国三大区域数据的比较研究》,《科技进步与对策》第5期。

赵渺希、唐子来,2010,《基于网络关联的长三角区域腹地划分》,《经济地理》第3期。

赵永亮、才国伟,2009,《市场潜力的边界效应与内外部市场一体化》,《经济研究》第7期。

赵永亮、徐勇、苏桂富,2008,《区际壁垒与贸易的边界效应》,《世界经济》第2期。

甄峰、王波、陈映雪,2012,《基于网络社会空间的中国城市网络特征——以新浪微博为例》,《地理学报》第8期。

周涛、高馨、罗家德,2022,《社会计算驱动的社会科学研究方法》,《社会学研究》第5期。

朱查松、王德、罗震东,2014,《中心性与控制力:长三角城市网络结构的组织特征及演化——企业联系的视角》,《城市规划学刊》第4期。

Afesorgbor, S. 2017, "Revisiting the Effect of Regional Integration on African Trade: Evidence from Meta-Analysis and Gravity Model." *The Journal of International Trade & Economic Development* 26(2).

Bairoch, P. 1991, *Cities and Economic Development: From the Dawn of History to the Present*, Chicago: University of Chicago Press.

Beaverstock, J., R. Smith & P. Taylor 2000, "World-City Network: A New Metageography?" *Annals of the Association of American Geographers* 90(1).

Bertinelli, L. & D. Black 2004, "Urbanization and Growth." *Journal of Urban Economics* 56(1).

Broker, J. 1984, *How Do International Trade Barriers Affect Interregional Trade? Regional and Industrial Theories*, Amsterdam: North-Holland Press.

Castells, M. 2000, *The Rise of the Network Society*, Oxford: Wiley-Blackwell Press.

Chen, Y. & L. Huang 2018, "A Scaling Approach to Evaluating the Distance Exponent of the Urban Gravity Model." *Chaos, Solitons & Fractals* 109.

Chen, Y. 2009, "Urban Gravity Model Based on Cross-Correlation Function and Fourier Analyses of Spatio-Temporal Process." *Chaos, Solitons & Fractals* 41(2).

Derudder, B. & F. Witlox 2005, "An Appraisal of the Use of Airline Data in Assessing the World City Network: A Research Note on Data." *Urban Studies* 42(13).

Engel, C. & J. Rogers 1994, *How Wide Is the Border?* Cambridge: National Bureau of Economic Research Cambridge.

Fang, C. & D. Yu 2017, "Urban Agglomeration: An Evolving Concept of An Emerging Phenomenon." *Landscape and Urban Planning* 162.

Fang, C. & D. Yu 2020, *China's Urban Agglomerations*, Singapore: Springer Singapore.

Fang, C., Q. Sheng & S. Li 2018, "Economic Links Between the Cities in North Anhui Based on Gravity Model." *Areal Research & Development* 37(2).

Gottmann, J. 1976, "Megalopolitan Systems Around the World." *Ekistics* 41(243).

Graham, S. & S. Marvin 1995, *Telecommunications and the City: Electronic Spaces, Urban Places*, London: Routledge.

Hall, P. & K. Pain 2006, *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*, London: Routledge.

Han, R., H. Cao & Z. Liu 2018, "Studying the Urban Hierarchical Pattern and Spatial Structure of China Using A Synthesized Gravity Model." *Science China Earth Sciences* 61(12).

Ibrahim, S. 2008, "A Study of Optimum Currency Area in East Asia: A Cluster Analysis." *Journal of Economic Integration* 23(4).

Kim, J. 2001, "Economic Integration of Major Industrialized Areas: An Empirical Tracking of the Continued Trend." *Technological Forecasting and Social Change* 67(2).

König, J. & R. Ohr 2013, "Different Efforts in European Economic Integration: Implications of the EU Index." *Journal of Common Market Studies* 51(6).

Lee, H., R. Owen & D. Mensbrugghe 2009, "Regional Integration in Asia and Its Effects on the EU and North America." *Journal of Asian Economics* 20(3).

Mahutga, M., X. Ma & D. Smith et al. 2010, "Economic Globalisation and the Structure of the World City System: The Case of Airline Passenger Data." *Urban Studies* 47(9).

Mayhew, B. & R. Levinger 1976, "Size and the Density of Interaction in Human Aggregates." *American Journal of Sociology* 82(1).

McCallum, J. 1995, "National Borders Matter: Canada-US Regional Trade Patterns." *The American Economic Review* 85(3).

McFarland, D., K. Lewis & A. Goldberg 2016, "Sociology in the Era of Big Data: The Ascent of Forensic Social Science." *The American Sociologist* 47(1).

Poncet, S. 2003, "Measuring Chinese Domestic and International Integration." *China Economic Review* 14(1).

Rossi, E., J. Beaverstock & P. Taylor 2007, "Transaction Links Through Cities: 'Decision Cities' and 'Service Cities' in Outsourcing by Leading Brazilian Firms." *Geoforum* 38(4).

Ruppert, E. 2013, "Rethinking Empirical Social Sciences." *Dialogues in Human Geography* 3(3).

Rutherford, J., A. Gillespie & R. Richardson 2004, "The Territoriality of Pan-European Telecommunications Backbone Networks." *Journal of Urban Technology* 11(3).

Smith, D. & M. Timberlake 2002, *Hierarchies of Dominance among World Cities: A Network Approach*, https://xueshu.baidu.com/usercenter/paper/show?paperid=941323ecde756d3bd80bad14a5dbeb9f&site=xueshu_se.

Storper, M. 2013, *Keys to the City: How Economics, Institutions, Social Interactions, and Politics Shape Development*, Princeton: Princeton University Press.

Taylor, P. 1999, "So-Called 'World Cities': The Evidential Structure Within A Literature." *Environment and Planning A* 31(11).

Van Nuffel, N., B. Derudder & F. Witlox 2010, "Even Important Connections Are Not Always Meaningful: On the Use of A Polarisation Measure in A Typology of European Cities in Air Transport Networks." *Tijdschrift voor economische en sociale geografie* 101(3).

Wenjie, L. 2022, "Evaluation of Urban-Rural Development Integration Level Based on Cluster Analysis." 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA).

Wilson, J. & T. Otsuki 2007, "Regional Integration in South Asia: What Role for Trade Facilitation?" World Bank Policy Research Working Paper.

Zhao, Y., G. Zhang & H. Zhao 2021, "Spatial Network Structures of Urban Agglomeration Based on the Improved Gravity Model: A Case Study in China's Two Urban Agglomerations." *Complexity*.

解密数字社会中的幸福密码

——不同互联网使用程度下社会信任感和社会公平感对主观幸福感的影响机制研究

郭媛媛　张　腾*

摘要：尽管已有大量研究探讨了主观幸福感的影响因素，但关于社会信任感、社会公平感与主观幸福感之间的关系，以及互联网使用程度如何调节这些关系的研究仍相对欠缺。对此，本文将社会信任感、社会公平感和主观幸福感依次设定为自变量、中介变量和因变量。研究对象为不同互联网使用程度的公众，数据来源于2021年中国社会调查数据（CGSS）。通过实证分析，本文有如下发现：首先，互联网使用程度对主观幸福感没有直接影响，受教育程度对主观幸福感呈现显著正向效应；其次，互联网使用程度越高，公众的社会公平感越低，不同省份的公众对社会公平感的感知具有一定差异；再次，性别差异也会影响社会公平感，女性的社会公平感更低；最后，社会信任感影响主观幸福感的内在机制是前者引起了公众社会公平感的变化。相对于其他群体，有时使用互联网的群体的社会信任感对社会公平感的影响最强，其社会公平感对主观幸福感的影响也最强；从不使用互联网的群体的社会信任感对主观幸福感的影响最强。据此，本文建议持续探究解救“数字难民”的路径，提升公众网络信息素养，建设绿色健康网络环境，提高公众主观幸福感。

关键词：互联网使用程度　信任感—公平感—幸福感　网络素养

一、引言

主观幸福感被认定为衡量社会保障水平和经济增长结果的关键指标，也是

* 郭媛媛，清华大学公共管理学院。张腾（通讯作者：ztdabao@163.com），清华大学公共管理学院。

国家发展和政策决策的重要目标(罗必良、洪炜杰、耿鹏鹏等,2021)。它不仅是评价国家或地区居民生活质量的重要标准(郭锦墉、王文锋、汤永红,2023),而且与个人心理健康、家庭和谐以及社会稳定密切相关。我国政府高度重视提高公众幸福感,并采取了一系列经济社会政策。在经济领域,我国制定了一系列政策,如"脱贫攻坚战"和"乡村振兴战略",旨在消除贫困,提升农村居民生活质量。国务院出台了相关文件,如《关于加强数字政府建设的指导意见》,进一步强调数字政府建设对数字经济、数字社会和数字生态的引领作用,促进经济社会高质量发展,不断增强人民群众的幸福感。在社会保障领域,国务院等相关部门发布了一系列政策文件,如《关于推进基本养老服务体系建设的意见》等,完善了养老、医疗、教育等领域的政策措施。在衣食住行育五大领域,全国医疗保险、保障性住房、全民义务教育等制度保障了公众基本生活条件的满足。然而,值得关注的是,改革开放 40 多年来,尽管经济增长显著,但公众的主观幸福感并未得到相应程度的提升。1990—2022 年,我国人均名义 GDP 增长了 52 倍,但与此同时公众的幸福感却从 7.3 分降至 5.59 分(郭锦墉、王文锋、汤永红,2023)。这种国家经济蓬勃发展与公众主观幸福感下降并存的现象,使我国政府面临一些挑战。

现有关于主观幸福感影响因素的研究主要分为两方面。首先是个体因素,包括健康状况、个人情绪和受教育程度等。多项研究表明,健康状况和个人情绪直接决定了个人的幸福感受(彭代彦、闫静,2014)。健康状况良好的人通常会有更高的主观幸福感,稳定的个人情绪、良好的受教育程度也是影响人们生活质量和主观幸福感的重要因素。其次是社会环境因素,如经济、文化和政治等。其中,经济发展水平、收入分配和消费水平等直接关系到人们的物质生活水平,以及对社会公平、主观幸福的感知(郭锦墉、王文锋、汤永红,2023);社会保障制度、社会信任感和社会公平感也对主观幸福感产生深远影响。在现代社会,特别是在互联网普及的时代,互联网使用程度的高低同样可能会影响公众的主观幸福感(丁雨桥,2023)。

尽管已有大量研究关注主观幸福感的影响因素,但社会信任、社会公平与主观幸福感之间的相互关系仍未得到充分研究。尤其值得注意的是,随着互联网的广泛应用,其在信息传播、交流方式以及价值观塑造等方面的作用为理解这三者的关系带来了新的可能性。互联网已经深度融入现代社会,极大地推动了知

识与资源的便利获取。然而,与此同时,网络暴力、虚假信息以及隐私泄露等问题也相应出现,可能削弱社会信任感、社会公平感,进而影响个体的主观幸福感。因此,如何理解互联网使用程度对社会信任感、社会公平感与主观幸福感之间关系的调和作用,是一个亟待探讨的问题。尽管互联网在日常生活中的影响日益显著,但在这一领域的研究仍相对匮乏。在此背景下,本文旨在探究不同互联网使用程度的公众主观幸福感的变化情况,以填补研究空白。本文将尝试回答以下研究问题:社会信任感、社会公平感分别如何影响主观幸福感?它们之间存在什么样的逻辑关系?互联网使用程度在社会信任感、社会公平感与主观幸福感之间起到怎样的调和作用?通过回答这些研究问题,本文将揭示社会信任感、社会公平感与主观幸福感的内在联系以及互联网使用程度在其中的调和作用,为相关政策制定者提供有益参考。

二、文献综述与理论框架

(一)理论基础

社会资本理论是一种解释社会结构、关系和互动对个人、集体行为影响的理论框架。它强调社会网络、规范和信任等要素在促进社会合作、降低社会成本及提高社会效率方面的重要作用。在本研究框架中,社会信任感被视为一种重要的社会资本,通过社会资本理论可以深入分析和解释社会信任感、社会公平感与主观幸福感之间的关系。从社会资本理论的角度看,社会信任感对主观幸福感的影响可以从多个层面加以理解。一方面,社会信任感直接影响主观幸福感。信任是社会资本的核心要素,当个体对他人和社会制度具有较高的信任时,他们更倾向于与他人建立联系、参与社会活动并积极投入社会事务。这种积极的社会参与和互动可以提高个体的幸福感,因为它们有助于满足个体对归属、尊重和自尊的需求。同时,信任还可以减轻个体在人际关系、社会交往中的紧张和担忧,从而使他们更容易体验到愉悦和幸福。另一方面,社会信任感还可以通过影响社会公平感来间接地影响主观幸福感。在社会资本理论中,信任被认为是一种有助于维护社会秩序、促进资源配置和实现社会目标的重要资源。当社会中

普遍存在较高的信任感时，人们更愿意遵守规则和法律，相信他人和制度能够确保公平地分配和对待。这种信任有助于提高个体对社会公平的感知，从而影响到他们的幸福感。在这一过程中，社会公平感作为中介变量起到了承上启下的作用。社会资本理论还强调了社会网络和规范在构建和维护信任中的关键作用。在本研究框架中，这一观点为分析社会信任感、社会公平感与主观幸福感之间的关系提供了重要的理论依据。具体来说，社会网络可以为个体提供信息、资源和支持，从而降低信任风险和成本，促进信任的形成和维持。

公平过程理论的核心观点在于，在对待公平问题上，人们不仅关注结果是否公平，还关注实现这一结果的过程是否公平。许多情况下，过程甚至比结果本身更重要。公平过程可带来积极的心理和社会效应，如提高人们的信任感、满意度和积极参与度。社会信任感在这一理论框架中发挥关键作用，因为它有助于确保公平过程的实现。高度的社会信任感意味着人们相信他人和制度会遵循公平的原则行事，这种信任感可以降低人们的防备心理，使他们更愿意积极参与社会活动并遵守社会规则。在公平过程理论中，社会公平感与主观幸福感密切相关。提高社会公平感往往会导致主观幸福感的提高，因为公平感让人们觉得权益受到保障、付出得到回报。在一个公平的社会中，人们更可能感受到尊重、认同和满足，从而提高幸福感。

综上所述，本研究框架结合了社会资本理论和公平过程理论，探讨了社会信任感、社会公平感与主观幸福感之间的关系。通过这一理论框架，研究者可以更好地理解社会信任感如何通过社会公平感影响主观幸福感，为相关政策制定者提供有益借鉴。

（二）主观幸福感

主观幸福感是一个多维概念，包括主观幸福感和心理幸福感等维度（Diener，Suh & Lucas et al.，1999）。主观幸福感指个体对自己生活满意度的主观评价，涉及社会关系、权利和福利等方面；心理幸福感关注个体的心理健康、情绪平衡和自我实现等方面。衡量主观幸福感的方法主要有两种：单项和多项。单项衡量方法通常使用一个简单的问题来评估个体的生活满意度（Cantril，1965）；多项衡量方法则包括多个方面的问题，如主观幸福感量表（Diener，Emmons &

Larsen et al.，1985）、心理幸福感量表（Ryff，1989）等。

众多研究揭示了主观幸福感与多种因素的关联，如个人特质（如性格、自尊等）（Diener，Suh & Lucas et al.，1999）、经济因素（如收入水平、经济增长等）（Easterlin，1974）、社会因素（如社会关系、社会支持等）（Helliwell & Putnam，2004）、心理因素（如心理健康、情绪调节等）（Lyubomirsky，King & Diener，2005）等，都与主观幸福感密切相关。此外，健康状况（Dolan，Peasgood & White，2008）、受教育水平（Oreopoulos & Salvanes，2011；Blanchflower & Oswald，2004）等也是影响主观幸福感的重要因素。良好的健康状况有助于提高个体的生活质量和满意度，也可能影响个体的心理和社会功能；较高的教育水平可能提高个体的社会地位、经济收入和职业满意度，进而提升主观幸福感。学者们还关注到社会信任感、社会公平感对主观幸福感的重要影响（Uslaner，2002；Alesina，Di Tella & MacCulloch，2004）。社会信任感体现为个体对他人及社会制度的信任程度。研究发现，社会信任感与主观幸福感存在显著正相关（Uslaner，2002）。信任有助于个体建立稳固的社会联系，获得更多的社会支持，并降低心理压力和焦虑，因为个体相信他人会遵守社会规范，不会损害自身利益（Putnam，2000）。社会公平感指个体对社会分配和社会制度公平性的评价。研究表明，社会公平感对主观幸福感具有重要影响（Alesina，Di Tella & MacCulloch，2004）。当个体感知到社会公平时，他们更有可能认为自己的努力会得到应有的回报，从而激发积极的心态和行为。此外，社会公平感还与社会凝聚力、社会稳定密切相关，这些因素也对主观幸福感具有积极作用（Wilkinson & Pickett，2009）。

（三）社会信任感

社会信任感是一种复杂的心理现象，涉及心理、文化、社会和制度等多个方面。目前，学界对社会信任的本质存在较大分歧，大多是从社会心理学、社会学和文化论等不同视角出发的观点解释。简而言之，社会心理学路径限于认知、情感和意向的预期，而社会学路径则从制度和社会变迁的角度研究态度以及态度与行为之间的关系。根据文化论观点，社会信任感是建立在习俗、传统和宗教基础上的文化现象（井世洁、杨宜音，2013）。

大量社会科学文献支持社会信任感对社会公平感的直接正向影响。比如，乌斯兰纳（E. Uslaner）提出，社会信任源于人们相信他人通常会做出公正的行为。他发现，高水平的社会信任感可以增强社区成员对社会公平感的认知，因为他们相信他人会公正地行事（Uslaner，2002）。罗思坦（B. Rothstein）等人的研究表明，社会信任感和社会公平感的关系是双向的。社会信任感不仅增加了社会公平感，公平的制度和政策也能够增强社会信任感（Rothstein & Stolle，2008）。这表明，社会公平感和社会信任感之间存在正向循环关系。此外，纳伊夫（M. Naef）等人的研究发现，在面对经济困境时，社会信任感在促进社会公平感方面起着核心作用（Naef & Schupp，2009）。他们发现，在资源相对有限时，社会信任感对增强社会公平感尤为重要。然而，并非所有研究结果都完全一致。例如，桑德斯科（K. Sønderskov）指出，在高度分化的社会中，社会信任感并不总是导致社会公平感（Sønderskov，2011）。然而，他的研究也表明，大多数情况下社会信任感确实可以增强社会公平感。井世洁等人认为，中国公众的社会信任感主要由政府信任感、社会组织信任感和社会信息信任感三方面构成，存在基于阶层和区域的结构性差异化特征（井世洁、杨宜音，2013）。基于上述分析，本文做出如下假设：

H1：社会信任感对社会公平感具有直接正向影响。

主观幸福感也被称为“生活满意度”（彭代彦、闫静，2014），社会信任感是影响个人生活满意度的重要因素之一，众多社会科学研究结果都支持这一结论。对世界价值观调查（WVS）中国部分数据的研究发现，相较于那些认为大多数人不可信且需特别小心谨慎的人，认为大多数人可以信赖的人报告的生活满意度较高（3.3%），而生活不满意的概率较低（0.9%）。赫利韦尔（J. Helliwell）等人的研究发现，社会信任感是主观幸福感的一个重要预测因子（Helliwell & Wang，2011）。他们收集了 34 个国家的大量数据，得出结论：高水平的社会信任感会增强个体的主观幸福感。此外，比约恩斯科夫（C. Bjørnskov）的研究提出，社会信任感不仅能提高人们的幸福感，还能提高整个社会的幸福水平（Bjørnskov，2008）。这是因为社会信任感能够减少社会冲突，促进人们的合作，创造更和谐的社会环境。郭锦墉等人指出，消费不平等通过降低人际信任和未来信心抑制了农村居民的主观幸福感（郭锦墉、王文锋、汤永红，2023）。哈德逊（J. Hud-

son)的研究表明,社会信任感和主观幸福感之间的关系可能受到其他因素的影响,如社会经济状况和政府效能(Hudson, 2006)。因此,社会信任感对主观幸福感的影响可能因不同的社会环境而异。基于上述分析,本文做出如下假设:

H2:社会信任感对主观幸福感具有直接正向影响。

(四)社会公平感

早期研究发现,社会公平感与主观幸福感之间存在显著的正相关关系(Alesina, Di Tella & MacCulloch, 2004)。学者们对多个国家的比较研究发现,社会公平感较高的国家其居民普遍具有较高的主观幸福感(Bjørnskov, Dreher & Fischer, 2007)。这可以从几个方面加以解释。

社会公平感对个体心理的影响至关重要。在公平的社会环境中,个体感受到自己的权益得到尊重,社会资源分配相对均衡,这使得个体的基本需求、价值追求更容易得到满足,从而提高主观幸福感(Brockner, Chen & Mannix et al., 2003);相反,当个体感知到社会不公,他们可能对未来抱有悲观情绪,担忧自身利益受损,从而导致主观幸福感降低(Rawls, 1971)。在公平的社会环境中,个体更容易建立互相信任的社会关系,形成更为稳固的社会支持网络(Wilkinson & Pickett, 2009)。这种社会支持对主观幸福感的提升具有重要作用,因为它能帮助个体应对压力,提高心理适应能力(Cohen & Wills, 1985)。此外,社会公平感还与政治参与、公民责任感相关。在公平的社会环境中,个体更有可能参与社会活动、履行公民责任,这些行为对个体的幸福感产生积极影响(Putnam, 2000);相反,在不公平的社会环境中,个体可能会感到愤怒和失望,产生消极的社会行为,从而降低主观幸福感(Gurr, 1970)。因此,本文提出如下假设:社会公平感与主观幸福感之间存在正向相关关系。这种社会支持对主观幸福感的提升具有重要作用,因为它能帮助个体应对压力,提高心理适应能力(Cohen & Wills, 1985)。此外,社会公平感还与政治参与和公民责任感相关。在公平的社会环境中,个体更有可能参与社会活动,履行公民责任,这些行为对个体的幸福感产生积极影响(Putnam, 2000)。相反,在不公平的社会环境中,个体可能会感到愤怒和失望,产生消极的社会行为,如罢工、抗议等,从而降低主观幸福感(Gurr, 1970)。基于上述分析,本文做出如下假设:

H3:社会公平感对主观幸福感具有直接正向影响。

（五）互联网使用程度

互联网使用对居民幸福感的影响是一个复杂的问题,国外部分学者从社会学、心理学角度展开了相关研究。目前,研究中存在两种对立的观点,即互联网信息抑郁论和互联网信息动员论。互联网信息抑郁论认为,面对海量信息时,个体容易产生心理压力和不适感,这种现象在数字时代尤为明显。互联网、社交媒体和其他数字平台上的信息爆炸式增长导致信息过载,使人们容易产生思维混乱、注意力分散、抑郁等问题,从而影响主观幸福感。同时,过度使用社交媒体可能导致焦虑、自卑感,与他人的“完美”生活进行比较使个体感到不满足和沮丧,进一步降低主观幸福感。在大量信息面前,人们的注意力难以集中,决策疲劳也会随之产生。此外,过度依赖网络寻求心理慰藉可能导致网络成瘾,影响日常生活和社交功能,进一步削弱主观幸福感(Robinson, 1976)。其中,互联网时间替代假说认为,过度使用互联网可能导致社会互动和交往时间的减少,进而影响主观幸福感。由于人际关系是主观幸福感的重要组成部分,因此其负面影响将对主观幸福感产生不良后果。

互联网信息动员论与互联网抑郁论相反,认为互联网和数字技术在不同使用场景下可以促进人际互动和联系,从而产生积极影响,提升公众主观幸福感。网络平台使人们能便捷地与他人建立联系,跨越地域和文化障碍实现全球信息共享。互联网的发展未必导致现实生活中社会互动的减少,反而可以帮助人们协作、组织活动和分享经验。

丁雨桥发现,中老年人使用互联网的频率越高,使用的工具越多元、用途越广泛,他们对生活的满意度也相对越高,这说明互联网的使用对于提高中老年人的主观幸福感具有一定促进作用(丁雨桥,2022)。此外,研究表明,中老年人使用互联网的程度越高、家庭情况越好、身体越健康,对提高他们的主观幸福感越有利。有研究发现,互联网使用可以增强社区的社会联系,从而提高社会信任感。这可能是因为互联网提供了丰富的信息和交流平台,使人们能够更容易地获取信息、分享经验、建立和维持社交关系。

根据 CGSS 2021 问卷原文,本文将互联网使用程度划分为五种类型,而本

文主要研究“从不使用”“有时使用”和“频繁使用”这三种类型。通过对比这三类互联网使用者,研究人员可以揭示互联网使用对现实生活中社会互动和主观幸福感的具体影响。此外,这种分类也有助于验证互联网时间替代假说和互联网信息动员论的适用性,并进一步探讨如何在网络与现实世界中找到平衡。基于上述分析,本文做出如下假设:

H4:频繁使用互联网的群体,其社会信任感对社会公平感的影响相对其他群体更显著。

H5:频繁使用互联网的群体,其社会信任感对主观幸福感的影响相对其他群体更显著。

(六)理论框架

根据上述分析,本文构建如下理论框架(图 1):

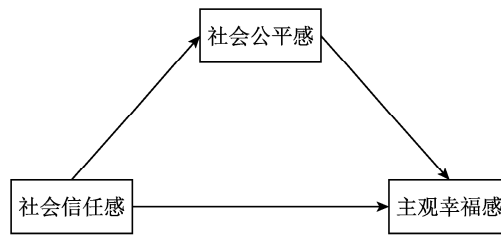


图 1　信任—公平—幸福三因素模型

三、研究方法

(一)数据来源

本文采用中国综合社会调查 CGSS 2021 的数据,为确保数据质量和准确性,进行了数据清洗和整理。在处理数据时,删除了缺失值样本,以提高分析结果的可靠性。通过对数据的严谨处理,本文旨在更准确地探讨在不同互联网使用程度下,社会信任感和社会公平感如何影响主观幸福感,为政策制定者和社会各界提供有针对性的建议,以提高公众主观幸福感。

（二）受访者基本情况

受访者基本情况如表 1：

表 1　　受访者基本情况

指标		数量	百分比	有效百分比	累计百分比
性别	男	2452	45.8	45.8	45.8
	女	2903	54.2	54.2	100.0
	总计	5355	100.0	100.0	
民族	汉族	4955	92.5	92.5	92.5
	蒙古族	10	0.2	0.2	92.7
	满族	22	0.4	0.4	93.1
	回族	119	2.2	2.2	95.4
	藏族	4	0.1	0.1	95.4
	壮族	100	1.9	1.9	97.3
	维吾尔族	145	2.7	2.7	100.0
	总计	5355	100.0	100.0	
宗教信仰	无宗教信仰	4952	92.5	92.5	92.5
	佛教	180	3.4	3.4	95.8
	道教	15	0.3	0.3	96.1
	民间信仰	17	0.3	0.3	96.4
	伊斯兰教	102	1.9	1.9	98.3
	天主教	11	0.2	0.2	98.5
	基督教	73	1.4	1.4	99.9
	其他	5	0.1	0.1	100.0
	总计	5355	100.0	100.0	
受教育程度	没受过任何教育	559	10.4	10.4	10.4
	扫盲班	28	0.5	0.5	11.0
	小学	1174	21.9	21.9	32.9
	初中	1509	28.2	28.2	61.1
	职业高中	89	1.7	1.7	62.7

（续表）

指标		数量	百分比	有效百分比	累计百分比
受教育程度	普通高中	628	11.7	11.7	74.5
	中专	238	4.4	4.4	78.9
	技校	21	0.4	0.4	79.3
	大学专科（成人高等教育）	175	3.3	3.3	82.6
	大学专科（正规高等教育）	246	4.6	4.6	87.2
	大学本科（成人高等教育）	135	2.5	2.5	89.7
	大学本科（正规高等教育）	478	8.9	8.9	98.6
	研究生及以上	75	1.4	1.4	100.0
	总计	5355	100.0	100.0	
地区	安徽省	290	5.4	5.4	5.4
	北京市	434	8.1	8.1	13.5
	福建省	213	4.0	4.0	17.5
	甘肃省	193	3.6	3.6	21.1
	广西壮族自治区	266	5.0	5.0	26.1
	河北省	205	3.8	3.8	29.9
	河南省	284	5.3	5.3	35.2
	湖北省	387	7.2	7.2	42.4
	湖南省	324	6.1	6.1	48.5
	江苏省	338	6.3	6.3	54.8
	江西省	333	6.2	6.2	61.0
	辽宁省	232	4.3	4.3	65.3
	内蒙古自治区	64	1.2	1.2	66.5
	宁夏回族自治区	73	1.4	1.4	67.9
	山东省	390	7.3	7.3	75.2
	山西省	203	3.8	3.8	79.0
	陕西省	251	4.7	4.7	83.7
	浙江省	312	5.8	5.8	89.5
	重庆市	563	10.5	10.5	100.0
	总计	5355	100.0	100.0	

从表 1 可知，就受访者性别而言，女性（54.2%）的比例略高于男性（45.8%）。在民族构成上，汉族占总人数的 92.5%，维吾尔族和壮族分别占 2.7%和 1.9%。关于宗教信仰，无宗教信仰者占多数，比例为 92.5%，佛教、基督教和伊斯兰教的信徒分别占 3.4%、1.4%和 1.9%。在受教育程度分布上，61.1%的人具有初中及以下的教育水平，而大学及以上学历的人占总人数的 10.3%。在地域分布方面，重庆市的占比最高（10.5%），其次是北京市（8.1%）和山东省（7.3%）。

（三）描述性统计

描述性统计分析结果如表 2：

表 2　　描述性统计分析结果

指标	最小值	最大值	平均值	标准差	偏度		峰度	
					统计	标准误差	统计	标准误差
互联网使用程度	1	5	3.33	1.670	-0.414	0.033	-1.520	0.067
社会公平感	1	5	3.67	0.993	-0.855	0.033	0.187	0.067
社会信任感	1	5	3.46	0.975	-0.769	0.033	-0.001	0.067
公众幸福感	1	5	3.99	0.819	-1.064	0.033	1.817	0.067

本文使用 SPSS 23.0 统计软件计算了 4 个观察变量的最小值、最大值、平均值、标准差、偏度和峰度。从最小值和最大值来看，所有观察变量的最小值和最大值都在合理取值范围内。平均值反映了一般水平，标准差反映了样本之间的差异程度。通过描述性统计可以发现，数据中没有异常值。一般用偏度和峰度来验证数据的正态分布：当数据服从绝对正态分布时，偏度和峰度都等于 0；但一般数据很难严格满足正态性条件，如果峰度绝对值小于 10 并且偏度绝对值小于 3，则说明数据虽然不是绝对正态，但可接受为近似正态分布。从表 2 可知：本样本峰度的绝对值全部小于 10，最大绝对值为 1.817；偏度的绝对值全部小于 3，最大绝对值为 1.064。该结果表明，数据可接受为近似正态分布。

四、结果与讨论

本文将应用场景分为三个群体:第一个场景是从不使用互联网的群体,即在5级李克特量表中选择选项1的公众;第二个场景是中等程度使用互联网(有时使用互联网)的群体,即在5级李克特量表中选择选项3的公众;第三个场景是频繁使用互联网的群体,即在5级李克特量表中选择选项5的公众。

(一) 二元 Logit 回归分析

1. 对主观幸福感的二元 Logit 回归分析

将 Provinces(省份)、Min(民族)、Educ(受教育程度)、Gender(性别)、Income(收入)、Religion(宗教)、IUD(互联网使用程度)作为自变量,将主观幸福感(SW)作为因变量,进行二元 Logit 回归分析(表3、表4)。

表3　　二元 Logit 回归模型似然比检验结果

模型	-2倍对数似然值	卡方值	*df*	*p*	AIC 值	BIC 值
截距	1021.595					
最终模型	1001.999	19.596	7	0.007	1017.999	1057.066

从表3可知,此处模型检验的原定假设为:是否放入自变量(Provinces、Min、Educ、Gender、Income、Religion、IUD)两种情况时模型质量均一样;$p<0.05$,拒绝原假设,说明本次构建模型时放入的自变量具有有效性,本次模型构建有意义。

表4　　二元 Logit 回归分析结果汇总

项	回归系数	标准误	*z* 值	Wald χ^2	*p* 值	OR 值	OR 值95% CI
Provinces	-0.026	0.016	-1.631	2.662	0.103	0.974	0.943—1.005
Min	-0.096	0.044	-2.189	4.794	0.029	0.909	0.834—0.990
Educ	0.062	0.030	2.048	4.195	0.041	1.064	1.003—1.129
Gender	-0.022	0.181	-0.123	0.015	0.902	0.978	0.686—1.395
Income	0.002	0.001	1.520	2.310	0.129	1.002	0.999—1.004

（续表）

项	回归系数	标准误	z 值	Wald χ^2	p 值	OR 值	OR 值 95% CI
Religion	-0.006	0.022	-0.268	0.072	0.789	0.994	0.952—1.038
IUD	0.117	0.184	0.638	0.407	0.523	1.124	0.784—1.612
截距	1.399	0.418	3.345	11.192	0.001	4.049	1.785—9.187
因变量：SW	McFadden R^2：0.019		Cox & Snell R^2：0.020		Nagelkerke R^2：0.031		

从表 4 可知，Provinces、Min、Educ、Gender、Income、Religion、IUD 可以解释 SW 的 0.02 变化原因。模型公式为：

$\ln(p/1-p) = 1.399-0.026 * \text{Provinces}-0.096 * \text{Min}+0.062 * \text{Educ}-0.022 * \text{Gender}+0.002 * \text{Income}-0.006 * \text{Religion}+0.117 * \text{IUD}$

其中 p 代表 SW 为 1 的概率，$1-p$ 代表 SW 为 0 的概率。

受教育程度对主观幸福感呈现显著的正相关性，这可能源于教育为个体带来的多维度积极效应。具体而言，较高的受教育程度可能提升个体的数字信息接入、获取和利用能力，提高社会经济地位，扩大职业机会，从而促进其生活质量的提升，进一步促成更高的主观幸福感（$z=2.048$，$p<0.05$）。教育还可能提高个体的问题解决能力和抗压技巧，这些均为提升主观幸福感的重要驱动力。另外，研究结果显示，省份、性别、收入、地区以及互联网使用程度等因素对 SW 并无显著影响。

2. 对社会公平感的二元 Logit 回归分析

将 Provinces、Min、Educ、Gender、Income、Religion、IUD 作为自变量，将 SF（社会公平感）作为因变量，进行二元 Logit 回归分析（表 5、表 6）。

表 5　　二元 Logit 回归模型似然比检验结果

模型	-2 倍对数似然值	卡方值	df	p	AIC 值	BIC 值
截距	1331.700					
最终模型	1300.949	30.750	7	0.000	1316.949	1356.017

从表 5 可知，此处模型检验的原定假设为：是否放入自变量（Provinces、Min、Educ、Gender、Income、Religion、IUD）两种情况时模型质量均一样；$p<0.05$，拒绝原假设，说明本次构建模型时放入的自变量具有有效性，本次模型构建有

意义。

表 6　　二元 Logit 回归分析结果汇总

项	回归系数	标准误	z 值	Wald χ^2	p 值	OR 值	OR 值 95% CI
Provinces	-0.048	0.014	-3.551	12.607	0.000	0.953	0.928—0.979
Min	-0.019	0.041	-0.470	0.221	0.639	0.981	0.905—1.063
Educ	0.002	0.023	0.101	0.010	0.920	1.002	0.958—1.049
Gender	-0.415	0.152	-2.728	7.442	0.006	0.660	0.490—0.890
Income	0.001	0.001	1.392	1.937	0.164	1.001	1.000—1.003
Religion	-0.004	0.019	-0.189	0.036	0.850	0.996	0.960—1.034
IUD	-0.367	0.159	-2.303	5.305	0.021	0.693	0.507—0.947
截距	1.788	0.358	4.993	24.927	0.000	5.975	2.962—12.053
因变量:SF	McFadden R^2:0.023		Cox & Snell R^2:0.031		Nagelkerke R^2:0.042		

将 Provinces、Min、Educ、Gender、Income、Religion、IUD 作为自变量,将 SF 作为因变量进行二元 Logit 回归分析。从表 6 可以看出,Provinces、Min、Educ、Gender、Income、Religion、IUD 可以解释 SF 的 0.02 变化原因。模型公式为:

$$\ln(p/1-p) = 1.788-0.048 * \text{Provinces}-0.019 * \text{Min}+0.002 * \text{Educ}-0.415 * \text{Gender}+0.001 * \text{Income}-0.004 * \text{Religion}-0.367 * \text{IUD}$$

其中 p 代表 SF 为 1 的概率,$1-p$ 代表 SF 为 0 的概率。

Provinces($z=-3.551, p<0.001$)、Gender($z=-2.728, p<0.01$)、IUD($z=-2.303, p<0.05$)会对 SF 产生显著的负向影响关系;但是 Min、Educ、Income、Religion 并不会对 SF 产生影响关系。也即,该结果具有三个结论。一是互联网使用程度越高,公众的社会公平感越低。互联网的普及和使用给人们提供了前所未有的信息获取途径,丰富的信息量反之又可以提升个体对社会公平的认知,形成反馈循环(feedback loop)。然而,该过程可能带来不确定的影响。频繁使用互联网的人群的社会公平感较低,这可能是因为信息爆炸时代社会不公平的事件传播速度更快,从而扭曲了个体对社会公平的总体评估。二是不同地区的公众对社会公平感的感知具有一定差异,地域差异发挥了重要作用。不同地区具有不同的经济、文化和政治背景,这些元素在很大程度上塑造了居民的价值观

和期望,从而影响了他们对社会公平的感知。三是性别差异会影响社会公平感,女性的社会公平感更低。社会角色理论(social role theory)认为,男性和女性在社会中所扮演的角色可能对他们的期望和认知产生影响。

(二)不同互联网使用程度的中介效应分析

1. 从不使用互联网

从不使用互联网群体的分析结果见表 7。

从表 7 可知,中介效应分析共涉及三个模型,分别如下:

A36 = 3.439−0.033 * Provinces−0.117 * Gender−0.026 * Min−0.001 * Religion+0.012 * Educ+0.001 * Income+0.254 * A33

A35 = 3.070−0.019 * Provinces−0.197 * Gender−0.003 * Min−0.019 * Religion−0.008 * Educ+0.001 * Income+0.282 * A33

A36 = 2.364−0.026 * Provinces−0.048 * Gender−0.025 * Min+0.006 * Religion+0.015 * Educ+0.000 * Income+0.155 * A33+0.350 * A35

社会信任感(A33)显著正向影响社会公平感(A35)(β = 0.282, p<0.01),表明从不使用互联网的群体其社会信任感越高,社会公平感也越高,H1 成立。这一发现支持了英格哈特(R. Inglehart)的观点,他指出社会信任感是一个重要的心理和社会资源,有助于增强个体对社会公平性的认同(Inglehart, 1997)。当个体在社会中感受到信任时,他们更容易相信社会制度是公平的,从而提高幸福感。相较于互联网用户,非互联网用户接触的信息量相对有限,这可能导致他们的社会公平感评估受到较少的消极信息影响。互联网上充斥着各种关于社会不公平的报道,这可能在一定程度上扭曲了互联网用户对社会公平性的评价;相比之下,非互联网用户可能对社会公平性持有更积极的观念。此外,非互联网用户可能更倾向于在现实生活中建立和维护社会联系。这种人际交往的亲密度和频率可能会加强社区的凝聚力,并形成更强的社会信任感。同时,这种紧密的社区环境也可能促进公平的社会规则和习俗的形成,从而提高社区成员对社会公平感的评估。

社会信任感(A33)和社会公平感(A35)均显著正向影响主观幸福感(A36)(分别为 β = 0.155, p<0.01; β = 0.350, p<0.01)。这意味着在从不使用互联网的

表 7　　中介作用分析结果　　N = 339

	A36					A35					A36				
	B	标准误	*t*	*p*	*β*	*B*	标准误	*t*	*p*	*β*	*B*	标准误	*t*	*p*	*β*
常数	3.439**	0.358	9.598	0.000	—	3.070**	0.392	7.833	0.000	—	2.364**	0.361	6.552	0.000	—
Provinces	−0.033**	0.011	−2.990	0.003	−0.156	−0.019	0.012	−1.550	0.122	−0.081	−0.026*	0.010	−2.581	0.010	−0.125
Gender	−0.117	0.122	−0.957	0.339	−0.052	−0.197	0.133	−1.479	0.140	−0.080	−0.048	0.113	−0.421	0.674	−0.021
Min	−0.026	0.027	−0.942	0.347	−0.049	−0.003	0.030	−0.089	0.929	−0.005	−0.025	0.025	−0.981	0.327	−0.047
Religion	−0.001	0.014	−0.072	0.943	−0.004	−0.019	0.015	−1.245	0.214	−0.065	0.006	0.013	0.436	0.663	0.021
Educ	0.012	0.038	0.314	0.754	0.017	−0.008	0.041	−0.195	0.846	−0.011	0.015	0.035	0.420	0.675	0.021
Income	0.001	0.000	1.545	0.123	0.081	0.001	0.000	1.816	0.070	0.095	0.000	0.000	0.914	0.361	0.044
A33	0.254**	0.050	5.075	0.000	0.265	0.282**	0.055	5.166	0.000	0.271	0.155**	0.048	3.221	0.001	0.162
A35											0.350**	0.046	7.526	0.000	0.380
R^2	0.120					0.109					0.249				
调整 R^2	0.101					0.090					0.231				
F 值	$F(7331)=6.444, p=0.000$					$F(7331)=5.787, p=0.000$					$F(8330)=13.667, p=0.000$				

注：* $p<0.05$，** $p<0.01$。下同。

群体中,社会信任感和社会公平感越高,主观幸福感也越高,H2、H3 成立。这一结果与乌斯兰纳的观点相符,即社会信任感、社会公平感是主观幸福感的关键决定因素(Uslaner, 2002)。此结果也与普特南(R. Putnam)的研究结果一致,他认为社会资本的积累对提高主观幸福感具有积极作用(Putnam, 2000)。卡瓦奇(I. Kawachi)等人也认为,社会信任感、社会公平感可以提高个体的心理健康水平,从而提高主观幸福感(Kawachi, Kennedy & Lochner et al., 1997)。社会信任感可以降低心理压力,减少担忧和恐惧,增强个体对未来的乐观预期;社会公平感可以满足个体对公平和正义的需求,降低社会不满和抱怨,从而提高主观幸福感。进一步而言,社会资本理论阐述了社会信任感、社会公平感作为社会资本的重要组成部分,对社会和谐稳定产生积极影响,从而间接提升个体的主观幸福感。感受到社会公正待遇的个体,会增强对社会的认同和归属感,降低心理压力,从而提升幸福感。此外,社会公平感、社会信任感可能通过其他机制影响主观幸福感。例如,高水平的社会信任感可以促进社会合作,降低社会交易成本,提高社会效率;强烈的社会公平感能降低社会矛盾和冲突,提高社会稳定性:两者都可能通过提升社会和个体的福利水平来提高主观幸福感。

从调整后的决定系数(R^2)来看,模型 3 的解释力最强(0.231),表明主观幸福感(A36)受到社会信任感(A33)、社会公平感(A35)以及其他控制变量的综合影响。

表 8　　中介作用检验结果汇总

项	c 总效应	a	b	a * b 中介效应值	a * b Boot SE	a * b *z* 值	a * b *p* 值	a * b 95% Boot CI	c′ 直接效应	检验结论
A33→A35→A36	0.254**	0.282**	0.350**	0.099	0.030	3.241	0.001	0.052—0.170	0.155**	部分中介

表 9　　中介作用效应量结果汇总

项	检验结论	c 总效应	a * b 中介效应	c′ 直接效应	效应占比计算公式	效应占比
A33→A35→A36	部分中介	0.254	0.099	0.155	a * b/c	38.958

在表7中，A33（社会信任感）对A35（社会公平感）的影响（a路径）以及A35对A36（主观幸福感）的影响（b路径）均为显著；同时，社会信任感（A33）通过社会公平感（A35）对主观幸福感（A36）的间接效应（a＊b路径）也非常显著（表8、表9）。这表明，社会公平感（A35）在社会信任感（A33）与主观幸福感（A36）之间具有部分中介作用。此结论与普特南、乌斯兰纳的研究观点相符。在表9中，社会信任感（A33）通过社会公平感（A35）对主观幸福感（A36）的间接效应占总效应的比例为38.958%，这意味着A35（社会公平感）在A33（社会信任感）与A36（主观幸福感）的关系中起到了显著的中介作用，证实了卡瓦奇、英格哈特的观点。研究结果表明，社会信任感对主观幸福感的影响的内在机制，是通过引起社会公平感的变化而实现的。社会资本理论及公平过程理论为理解社会信任感对主观幸福感影响的内在机制提供了理论基础。在社会资本理论视角下，社会信任感被视作社会资本的核心要素，其中社会资本被定义为社会网络中内嵌的信息、知识、信任及规范，这些资本能为个体或群体带来多样性收益。社会信任感可以有效促成社会关系的建立与维系，提供社会支援，降低交易成本，增进合作精神，从而通过提升社会资本来增强主观幸福感。公平过程理论强调了社会公平感对个人主观幸福感的显著影响，无论是对结果的公平感还是对过程的公平感，都对个体的主观幸福感产生了深远影响。当社会公平的感知增强时，社会满意度和主观幸福感相应提升。随着社会信任感的增强，社会资本得到提升，个体对社会公平的感知增强，因而主观幸福感也在增强。

2. 有时使用互联网

有时使用互联网群体的分析结果见表10。

从表10可知，中介效应分析共涉及三个模型，分别如下：

A36=2.798−0.012＊Provinces+0.147＊Gender−0.029＊Min+0.015＊Religion+0.033＊Educ+0.228＊A33

A35=1.870−0.003＊Provinces−0.076＊Gender+0.036＊Min+0.005＊Religion+0.043＊Educ+0.383＊A33

A36=2.377−0.011＊Provinces+0.164＊Gender−0.037＊Min+0.014＊Religion+0.024＊Educ+0.142＊A33+0.225＊A35

自变量A33（社会信任感）对A35（社会公平感）的显著正向影响（$B=0.383$，

表 10 **中介作用分析结果** N = 416

	A36					A35					A36				
	B	标准误	*t*	*p*	*β*	*B*	标准误	*t*	*p*	*β*	*B*	标准误	*t*	*p*	*β*
常数	2.798**	0.246	11.392	0.000	—	1.870**	0.288	6.484	0.000	—	2.377**	0.249	9.544	0.000	—
Provinces	−0.012	0.007	−1.597	0.111	−0.077	−0.003	0.009	−0.400	0.689	−0.019	−0.011	0.007	−1.544	0.123	−0.072
Gender	0.147	0.081	1.807	0.072	0.086	−0.076	0.095	−0.794	0.428	−0.036	0.164*	0.078	2.086	0.038	0.096
Min	−0.029	0.032	−0.881	0.379	−0.042	0.036	0.038	0.937	0.350	0.043	−0.037	0.031	−1.167	0.244	−0.054
Religion	0.015	0.011	1.438	0.151	0.069	0.005	0.013	0.411	0.681	0.019	0.014	0.010	1.377	0.169	0.063
Educ	0.033	0.019	1.753	0.080	0.084	0.043	0.022	1.915	0.056	0.088	0.024	0.019	1.286	0.199	0.060
A33	0.228**	0.038	5.919	0.000	0.280	0.383**	0.045	8.475	0.000	0.386	0.142**	0.040	3.515	0.000	0.174
A35											0.225**	0.041	5.533	0.000	0.274
R^2	0.106					0.171					0.168				
调整 R^2	0.092					0.159					0.154				
F 值	$F(6409) = 8.047, p = 0.000$					$F(6409) = 14.040, p = 0.000$					$F(7408) = 11.771, p = 0.000$				

$p<0.01$）表明（表 10），当个体在社会中感受到更高的信任水平时，倾向于认为社会更加公平。在这种情况下，社会信任感作为一种资源，有助于缓解潜在的不公平现象，从而提高个体对社会公平的感知。自变量 A33（社会信任感）对 A36（主观幸福感）的显著正向影响（$B=0.142$，$p<0.01$）表明（表 10），社会信任感对个体的主观幸福感有积极作用。普特南的研究证实了这一观点，他强调社会信任感在形成社会凝聚力和主观幸福感方面发挥着重要作用。社会信任感使人们能够建立更紧密的关系，加强彼此之间的合作，从而提高个体的幸福感。中介变量 A35（社会公平感）对 A36（主观幸福感）的显著正向影响（$B=0.225$，$p<0.01$）表明（表 11），社会公平感对个体的幸福感起到关键作用。这与罗尔斯（J. Rawls）的正义观相符，他认为公平是社会正义的核心原则，公平的社会环境对于个体主观幸福感的形成至关重要（Rawls，1971）。当个体感受到社会公平时，他们更容易认为自己在社会中受到尊重和公平对待，从而提高幸福感。在控制变量方面，性别（Gender）对 A36（主观幸福感）的显著正向影响（$B=0.164$，$p<0.05$）表明，不同性别在幸福感方面存在差异。

表 11　　**中介作用检验结果汇总**

项	c	a	b	a * b	a * b	a * b	a * b	a * b	c′	检验结论
	总效应			中介效应值	Boot SE	z 值	p 值	95% Boot CI	直接效应	
A33→A35→A36	0.228**	0.383**	0.225**	0.086	0.026	3.274	0.001	0.058—0.160	0.142**	部分中介

表 12　　**中介作用效应量结果汇总**

项	检验结论	c	a * b	c′	效应占比计算公式	效应占比
		总效应	中介效应	直接效应		
A33→A35→A36	部分中介	0.228	0.086	0.142	a * b/c	37.831

表 13　　**效应值 95%置信区间**

项	c	a	b	a * b	c′
	总效应			中介效应	直接效应
A33→A35→A36	0.152—0.303	0.294—0.471	0.145—0.305	0.058—0.160	0.063—0.220

注：a * b 为 95% bootstrap CI。下同。

在有时使用互联网的情境下，分析结果显示，A33（社会信任感）通过 A35（社会公平感）对 A36（主观幸福感）产生部分中介作用（表 12、表 13）。这些研究结果可以解释为什么在有时使用互联网的情境下，社会信任感会影响主观幸福感。首先，当个体对周围环境的信任感较高时，他们更容易认为社会是公平的。这种信任感可能来自他们在互联网上与他人的互动，这些互动可能增强了他们对社会的信任。其次，社会公平感被认为是个体主观幸福感的一个重要决定因素（Kawachi，Kennedy & Lochner et al.，1997）。当个体认为社会公平时，他们可能更容易满足于自己的需求和目标，从而提高幸福感。综上所述，在有时使用互联网的情境下，社会信任感影响主观幸福感的内在机制，是通过引起社会公平感的变化达成的。这表明在这种情境下，提高社会信任感、社会公平感有助于提高个体的主观幸福感。

3. 频繁使用互联网

频繁使用互联网群体的分析结果见表 14。

从表 14 可知，中介效应分析共涉及三个模型，分别如下：

A36 = 3. 230 + 0. 050 * Gender + 0. 007 * Min − 0. 000 * Religion + 0. 019 * Educ − 0. 000 * Provinces + 0. 163 * A33

A35 = 2. 165 − 0. 031 * Gender + 0. 008 * Min + 0. 015 * Religion + 0. 016 * Educ + 0. 004 * Provinces + 0. 304 * A33

A36 = 2. 795 + 0. 057 * Gender + 0. 006 * Min − 0. 003 * Religion + 0. 016 * Educ − 0. 001 * Provinces + 0. 102 * A33 + 0. 201 * A35

自变量 A33（社会信任感）对 A35（社会公平感）的显著正向影响（B = 0. 304，p<0. 01）表明（表 14），对于频繁使用互联网的人来说，当社会信任感增加时，他们对社会公平感的评价也会提高。这再次验证了科尔曼的社会资本理论，即社会信任感与社会公平感之间的积极联系。自变量 A33（社会信任感）对 A36（主观幸福感）的显著正向影响（B = 0. 102，p<0. 01）表明（表 14），在频繁使用互联网的群体中，社会信任感对于个体幸福感的形成同样具有积极作用。这与普特南的观点相吻合，他强调社会信任感对于社会凝聚力和主观幸福感的重要性。中介变量 A35（社会公平感）对 A36（主观幸福感）的显著正向影响（B = 0. 201，p<0. 01）（表 15）证实，在频繁使用互联网的群体中，社会公平感对主观

表 14 中介作用分析结果 N = 2016

	A36					A35					A36				
	B	标准误	*t*	*p*	*β*	*B*	标准误	*t*	*p*	*β*	*B*	标准误	*t*	*p*	*β*
常数	3.230**	0.098	32.826	0.000	—	2.165**	0.119	18.220	0.000	—	2.795**	0.103	27.119	0.000	—
Gender	0.050	0.033	1.544	0.123	0.034	−0.031	0.039	−0.791	0.429	−0.017	0.057	0.032	1.788	0.074	0.038
Min	0.007	0.014	0.514	0.607	0.011	0.008	0.017	0.462	0.644	0.010	0.006	0.014	0.414	0.679	0.009
Religion	−0.000	0.006	−0.041	0.967	−0.001	0.015*	0.007	2.114	0.035	0.045	−0.003	0.006	−0.570	0.569	−0.012
Educ	0.019**	0.005	3.988	0.000	0.088	0.016**	0.006	2.710	0.007	0.058	0.016**	0.005	3.426	0.001	0.073
Provinces	−0.000	0.003	−0.087	0.931	−0.002	0.004	0.004	1.000	0.318	0.021	−0.001	0.003	−0.339	0.735	−0.007
A33	0.163**	0.017	9.740	0.000	0.212	0.304**	0.020	15.059	0.000	0.319	0.102**	0.017	5.947	0.000	0.133
A35											0.201**	0.018	11.201	0.000	0.250
R^2	0.054					0.107					0.110				
调整 R^2	0.052					0.105					0.107				
F 值	$F(62\,009) = 19.276, p = 0.000$					$F(62\,009) = 40.288, p = 0.000$					$F(72\,008) = 35.468, p = 0.000$				

幸福感的形成具有关键作用。这与罗尔斯的正义观一致,即公平是社会正义的基础,公平的社会环境有助于提高个体主观幸福感。对于控制变量,受教育程度(Educ)对 A36(主观幸福感)和 A35(社会公平感)均存在显著正向影响(分别为 $B=0.016$, $p<0.01$; $B=0.019$, $p<0.01$)。这表明,受教育程度较高的个体往往对社会公平感、主观幸福感有更高的评价。

表 15 中介作用检验结果汇总

项	c	a	b	a * b	a * b	a * b	a * b	a * b	c′	检验结论
	总效应			中介效应值	Boot SE	*z* 值	*p* 值	95% Boot CI	直接效应	
A33→A35→A36	0.163**	0.304**	0.201**	0.061	0.010	6.342	0.000	0.061—0.099	0.102**	部分中介

表 16 中介作用效应量结果汇总

项	检验结论	c	a * b	c′	效应占比计算公式	效应占比
		总效应	中介效应	直接效应		
A33→A35→A36	部分中介	0.163	0.061	0.102	a * b/c	37.492

表 17 效应值 95%置信区间

项	c	a	b	a * b	c′
	总效应			中介效应	直接效应
A33→A35→A36	0.130—0.196	0.264—0.344	0.166—0.236	0.061—0.099	0.068—0.135

表 15、表 16 总结了中介作用的检验结果。A33 通过 A35(互联网使用满意度)对 A36(主观幸福感)产生部分中介作用。表 15 显示,A33→A35 的路径系数(a)是 0.304($p<0.01$),A35→A36 的路径系数(b)是 0.201($p<0.01$),A33 直接影响 A36 的路径系数(c′)是 0.102($p<0.01$),中介效应值(a * b)是 0.061。通过自助法检验,中介效应的 95%置信区间为 0.061—0.099,说明中介效应是显著的。表 16 显示,A33 通过 A35 对 A36 产生的部分中介效应占总效应的比例为 37.492%。表 17 提供了效应值的 95%置信区间,A33 对 A36 的总效应值的置信区间为 0.130—0.196,A33 对 A35 的总效应值的置信区间为 0.264—0.344,A35 对 A36 的总效应值的置信区间为 0.166—0.236,中介效应值的置信

区间为 0. 061—0. 099，A33 直接影响 A36 的总效应值的置信区间为 0. 068—0. 135。

总之，本文得出如下结论：

第一，在三种互联网使用程度的群体中，社会信任感对社会公平感、主观幸福感均有显著影响，社会公平感对主观幸福感也有显著影响。这表明，无论互联网使用程度如何，社会信任感、社会公平感和主观幸福感之间均存在显著关系。然而，在从不使用互联网的群体中，社会信任感对社会公平感的解释力（R^2）为 0. 109；在有时使用互联网的群体中，该值为 0. 171；在频繁使用互联网的群体中，该值为 0. 107。这表明，在有时使用互联网的群体中，社会信任感对社会公平感的解释力较强。此外，在从不使用互联网的群体中，社会信任感对社会公平感的影响程度（β）为 0. 282；在有时使用互联网的群体中，该值为 0. 383；在频繁使用互联网的群体中，该值为 0. 304。这说明，在有时使用互联网的群体中，社会信任感对社会公平感的影响程度最高，H4 不成立。这可能是因为有时使用互联网的群体在获取信息、建立社交关系和评估社会环境方面拥有更多的选择和资源（Castells，2000）。互联网为他们提供了丰富的信息来源，使他们能够综合分析社会信任感和社会公平感，并将这两个因素与个体主观幸福感相联系（Diener，Oishi & Lucas，2003）。

第二，在三种互联网使用程度的群体中，社会信任感对主观幸福感均有显著影响。在从不使用互联网的群体中，社会信任感对主观幸福感的影响程度（β）为 0. 155；在有时使用互联网的群体中，该值为 0. 142；在频繁使用互联网的群体中，该值为 0. 102。这说明，在从不使用互联网的群体中，社会信任感对主观幸福感的影响程度最高，H5 不成立。对于从不使用互联网的群体，其社交范围主要在物理社区内，更多依赖于面对面的交往，使社会信任感对主观幸福感的影响更为显著。公平过程理论指出，个体对于环境公正的感受程度决定了其社会满意度，进而影响其主观幸福感。未接触互联网的群体通过直接的社会交往来判断公平性，因此社会信任感在他们的主观幸福感中占据了更大的比重。

第三，在三种互联网使用程度的群体中，社会公平感对主观幸福感均有显著影响。在从不使用互联网的群体中，社会公平感对主观幸福感的影响程度（β）为 0. 282；在有时使用互联网的群体中，该值为 0. 383；在频繁使用互联网的群体

中,该值为 0.304。这说明,在有时使用互联网的群体中,社会公平感对主观幸福感的影响程度最强。该现象可能源于该群体在信息获取、社会公平性判断及网络信息甄别等方面的优势。该群体可以通过互联网接触更多信息,同时仍能通过现实社交获取具有针对性的信息,从而形成全面而深入的社会公平性理解。与频繁使用互联网的群体相比,他们较少受到网络负面信息的冲击;与未接触互联网的群体相比,他们对社会公平性有更准确的判断。因此,社会公平感在此类群体中对主观幸福感的影响尤为显著。

第四,互联网使用程度的提高,降低了公众社会信任感对社会公平感的影响,以及社会信任感、社会公平感对主观幸福感的影响。即使个人对社会的信任感相对较高,他们也可能因为接触一些揭示社会不公的信息而对社会公平性产生质疑。这就可能解释为什么随着互联网使用程度的提高,社会信任感对社会公平感的影响减弱。

这些差异揭示了互联网使用程度可能会影响个体对社会信任感、社会公平感与主观幸福感之间关系的认知。不同程度的互联网使用可能会导致个体在信息获取、社交互动等方面的差异,进而影响他们对社会信任感、社会公平感与主观幸福感的关联程度的评估。

公众主观幸福感对一个国家经济社会的稳定性和发展的可持续性具有十分重要的影响。本文研究基于上述实证结论,提出以下政策建议:

第一,政府部门应对公众主观幸福感的影响机制有所关注并采取一些行动。首先要通过增加民生财政支出,补齐基本公共服务短板,满足公众多样化的公共服务需求,逐步增强公众对社会发展和幸福生活的信心;其次要提高群体教育水平,利用教育过程中产生的安全感与获得感对社会信任提供正向影响;最后要密切关注群体社会关系,通过发展社会中介组织,正确发扬群体效应,增强公众的人际信任和发展信心,确保社会信任感在提升个体主观幸福感中发挥更大作用。

第二,针对不同的互联网使用程度群体,政府应实施提升主观幸福感的差异化策略。首先,对频繁使用互联网的群体,政府部门应采取教育、正面引导等手段来提高这部分群体的信息甄别意识和理性判断能力,进而削弱误导性、欺骗性网络信息对社会信任和社会公平的危害。其次,对从不使用互联网的群体,政府应通过提高大众教育水平、宣传互联网媒介传播效应等方式,使该群体学会利用

互联网中的海量有利信息应对现实生活中的不确定性风险，提升自身的主观幸福感。最后，针对有时使用互联网的群体，政府需要采取均衡发展策略来提高这部分群体的主观幸福感，如：推广并强化数字素养教育；建立和完善信息公开平台，以便该群体更便捷地获取公正、准确的社会信息；注意维护和保护该群体的线下社区生活；鼓励并协助该群体在互联网上发挥更大的价值。

第三，针对女性社会公平感相对较低的现象，应注意以下方面：一是积极地从根源上打破性别刻板印象；二是加强对女性经济权益的保护；三是增加女性在社会决策中的占比和影响力；四是加强网络环境的建设和监管，保护女性免受网络骚扰。

五、结论

本文通过对不同互联网使用程度群体的对比分析，对主观幸福感、社会公平感、社会信任感之间的关系进行了探讨，得出了一系列重要结论：一是互联网使用程度并未对主观幸福感产生直接影响，否定了互联网信息抑郁论和互联网信息动员论；二是受教育程度对主观幸福感呈现显著正向效应；三是随着互联网使用程度的提高，公众对社会公平感的感知降低；四是不同地区的公众对社会公平感的感知存在显著差异，表明社会公平感受地域影响；五是性别差异影响对社会公平性的感知，女性对社会公平性的感知相对较低。进一步的分析揭示了社会信任感影响主观幸福感的内在机制，即通过引发公众对社会公平感的变化而达成。特别是，在有时使用互联网的群体中，社会信任感对社会公平感的影响最强，社会公平感对主观幸福感的影响也最强；而在从不使用互联网的群体中，社会信任感对主观幸福感的影响最强。

本文的创新点在于首次从互联网使用程度的角度，系统地探讨其与社会信任感、社会公平感以及主观幸福感之间的关系，既丰富了互联网使用对社会影响的研究成果，又为理解主观幸福感的影响因素提供了新的视角。本文结合社会资本理论与公平过程理论，揭示了社会信任感影响主观幸福感的内在机制，这种理论整合提供了一个更全面的框架来理解社会信任感、社会公平感和主观幸福感之间的复杂关系。总之，本文关注互联网使用程度对社会信任感、社会公平感

与主观幸福感关系的影响，并通过对比不同互联网使用程度群体的数据，揭示互联网使用对于不同群体感知这些变量的影响，以期为政策制定者提供有益参考。

参考文献

丁雨桥，2022，《互联网使用对中老年人生活幸福感的影响》，南京邮电大学硕士学位论文。

郭锦墉、王文锋、汤永红，2023，《消费不平等对农村居民主观幸福感的影响及作用机理研究》，《农林经济管理学报》第 3 期。

井世洁、杨宜音，2013，《转型期社会信任感的阶层与区域特征》，《社会科学》第 6 期。

罗必良、洪炜杰、耿鹏鹏等，2021，《赋权、强能、包容：在相对贫困治理中增进农民幸福感》，《管理世界》第 10 期。

彭代彦、闫静，2014，《社会信任感与生活满意度——基于世界价值观调查（WVS）中国部分数据的实证分析》，《当代经济研究》第 6 期。

Alesina, A., R. Di Tella & R. MacCulloch 2004, "Inequality and Happiness: Are Europeans and Americans Different?" *Journal of Public Economics* 88(9-10).

Bjørnskov, C. 2008, "Social Trust and Fractionalization: A Possible Reinterpretation." *European Sociological Review* 24(3).

Bjørnskov, C., A. Dreher & J. Fischer 2007, "The Bigger the Better? Evidence of the Effect of Government Size on Life Satisfaction Around the World." *Public Choice* 130.

Blanchflower, D. & A. Oswald 2004, "Well-Being over Time in Britain and the USA." *Public Economic* 88(7-8).

Brockner, J., Y. Chen & E. Mannix et al. 2003, "Culture and Procedural Fairness: When the Effects of What You Do Depend on How You Do It." *Administrative Science Quarterly* 45(1).

Cantril, H. 1965, *The Pattern of Human Concerns*, New Brunswick: Rutgers University Press.

Castells, M. 2000, *The Rise of the Network Society*, Oxford: Blackwell Publishers.

Cohen, S. & T. Wills 1985, "Stress, Social Support, and the Buffering Hypothesis." *Psychological Bulletin* 98(2).

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

Diener, E., R. Emmons & R. Larsen et al. 1985, "The Satisfaction With Life Scale." *Journal of Personality Assessment* 49(1).

Diener, E., S. Oishi & R. Lucas 2003, "Personality, Culture, and Subjective Well-Being: Emotional and Cognitive Evaluations of Life." *Annual Review of Psychology* 54.

Dolan, P., T. Peasgood & M. White 2008, "Do We Really Know What Makes Us Happy? A Review of the Economic Literature on the Factors Associated with Subjective Well-Being." *Journal of Economic Psychology* 29(1).

Easterlin, R. 1974, "Does Economic Growth Improve the Human Lot? Some Empirical Evidence." in P. David & M. Reder eds., *Nations and Households in Economic Growth*, Amsterdam: Academic Press.

Gurr, T. 1970, *Why Men Rebel*, Princeton: Princeton University Press.

Helliwell, J. & R. Putnam 2004, "The Social Context of Well-Being." https://pubmed.ncbi.nlm.nih.gov/15347534.

Helliwell, J. & S. Wang 2011, "Trust and Wellbeing." https://www.nber.org/papers/w15911.

Hudson, J. 2006, "Institutional Trust and Subjective Well-Being Across the EU." https://doi.org/10.1111/j.1467-6435.2006.00319.x.

Inglehart, R. 1997, *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*, Princeton: Princeton University Press.

Kawachi, I., B. Kennedy & K. Lochner et al. 1997, "Social Capital, Income Inequality, and Mortality." *American Journal of Public Health* 87(9).

Lyubomirsky, S., L. King & E. Diener 2005, "The Benefits of Frequent Positive Affect: Does Happiness Lead to Success?" *Psychological Bulletin* 131(6).

Naef, M. & J. Schupp 2009, "Measuring Trust: Experiments and Surveys in Contrast and Combination." https://www.iza.org/publications/dp/4087/measuring-trust-experiments-and-surveys-in-contrast-and-combination.

Oreopoulos, P. & K. Salvanes 2011, "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1).

Putnam, R. 2000, *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon & Schuster.

Rawls, J. 1971, *A Theory of Justice*, Cambridge: Belknap Press of Harvard University Press.

Robinson, M. 1976, "Public Affairs Television and the Growth of Political Malaise: the Case of the Selling of the Pentagon." *American Political Science Review* 70(2).

Rothstein, B. & D. Stolle 2008, "The State and Social Capital: An Institutional Theory of Generalized Trust." *Comparative Politics* 40(4).

Ryff, C. 1989, "Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being." *Journal of Personality and Social Psychology* 57(6).

Sønderskov, K. 2011, "Does Generalized Social Trust Lead to Associational Membership? Unravelling A Bowl of Well-Tossed Spaghetti." *European Sociological Review* 27(4).

Uslaner, E. 2002, *The Moral Foundations of Trust*, Cambridge: Cambridge University Press.

Wilkinson, R. & K. Pickett 2009, "Income Inequality and Social Dysfunction." *Annual Review of Sociology* 35.

数字时代安全科技价值报告*

陆菲菲　秦　昊　张友红　吕　鹏**

摘要：本文基于2023年全球数字社会的安全风险新形势和科技创新内在逻辑，对安全科技做出“公共品”的新定义，总结出其具有的压舱石与助燃剂双重价值，提出人工智能和安全科技将会是未来的两项通用技术。政企合作是安全科技成为公共品的核心驱动力，通过更多的公共力量投入，构建更好的安全科技实力和制度，采用数字化转型的“慢起飞模型”，新的安全技术将得到更好的发展，个人与社会为数字化转型所付出的代价也将更小。

关键词：数字时代　安全科技　AI安全　大模型安全　网络安全　数据安全

一、前言

安全已经成为数字时代最为重要的话题，我们正处于安全风险的质变时刻。随着数据爆炸式增长，数字化基础架构错综复杂，网络攻击面不断扩大，安全风险也呈现出新情况。尤其是在信息安全方面，近年针对数据的攻击、窃取、滥用等各种风险事件不断涌现，对经济、社会产生了巨大影响。

2023年人工智能（AI）大模型横空出世，大大推进了深度智能时代的到来，然而它对安全局面的影响同样可能是颠覆性的。最近引起全球关注的OpenAI公司因对AI安全的不同观点而导致的人员变动风波，很好地说明了业界对当前

* 本文系中国社会科学院2024年度实验室孵化专项资助“社会计算与数字社会前沿研究”（项目批准号：2024SYFH010）的阶段性研究成果。感谢中国社会科学院大学哲学院教授、中国社会科学院科学技术和社会研究中心主任段伟文，科技部“云计算+大数据”国家重点专项课题负责人、数秦科技有限公司执行总裁兼首席科学家崔伟，北京大学客座教授、蚂蚁集团首席技术安全官韦韬，浙江工业大学网络空间安全研究院院长宣琦等多位专家学者在本文写作过程中提出的宝贵观点和建议。

** 陆菲菲，蚂蚁集团安全实验室。秦昊，蚂蚁集团安全实验室。张友红，蚂蚁集团安全实验室。吕鹏，中国社会科学院大学数字中国研究院。

AI 安全高风险的态度，以及它在整个深度智能社会中投下的令人不安的阴影。因此，站在深度数字化的转折时刻，对安全科技的关注与投入已刻不容缓。

安全科技的价值，既体现为压舱石，更重要的是作为助燃剂：压舱石指的是守住技术的安全底线，防御可能存在的风险隐患，支撑和保障数字社会的稳定运行；助燃剂是提高技术的安全上限，降低技术运行的成本，让新技术得以规模化落地，推动产业焕发新机，让更多人从中受益。

数字化为何有安全风险？其实，风险自古至今都存在，只是因为产业运营、个人生活、国家组织的行为模式向数字化转型，风险也就相应地并轨转移。隐私数据泄露、商业勒索、基础设施瘫痪……放眼未来，有新科技的地方，就会有新的数字化安全风险，两者是本体和阴影一样不可分割的伴生关系。

我们在前段时间经过综合调研，同时结合多位专家学者提供的宝贵观点和建议，撰写了这份报告。本报告主要关注两大议题：第一，在安全风险质变时刻，更好地理解数字时代的安全风险新形势和科技创新的内在逻辑，提高社会对安全科技的认知、重视与参与度，这是本报告的核心议题；第二，安全科技行业存在多种需要迫切解决的问题，尤其是缺乏明确的生态标准来加快前沿技术的规模化落地，缺乏合理的市场化机制及生态联动方法推动安全科技健康发展，这也是本报告关注的关键问题。

本文的主要结论是安全科技将成为社会的公共品，是数字社会的优先事项之一。未来通用的技术有两个，一是人工智能，二是安全科技。人工智能作为最具颠覆性和战略性的关键核心技术，将在各行各业数字化的基础上进一步实现智能化，成为未来生产力的发动机；安全科技将作为方向盘，始终将各种新兴科技控制在向善的道路上，它保护的不只是若干行业、某个业务、几段数据，而是这个社会里的每一个人，是技术向善的关键保障。安全科技与人工智能技术并举，双方的融合创新与发展，将是我国强国战略中不可忽视的重要助推因素。因此，这两个领域，应该成为我们最关注、投入最大的领域。

我们希望通过这份安全科技价值的研究报告，能更加深刻地洞悉安全科技的未来发展之路，提高社会对安全科技的认知，实现安全科技与人工智能的并举、融合与创新，助力经济社会高质量发展。

二、数字社会面临的全新风险挑战

（一）数字化安全风险的三个历史阶段

数字化安全风险的历史与信息技术的进步保持同步，并非最新的产物。它长久以来处于公众雷达之外，本质原因是 20 世纪末 21 世纪初我国上网民众基数不高。但在数字化突飞猛进的今天，数字化安全风险日益引发关注。我们可以把数字化安全风险的演变划分为三个阶段。

第一阶段是互联网“上古”时期至 20 世纪 90 年代初，这是网络安全风险的萌芽时期。事实上，前互联网时代就已经有了网络攻击，因为保护位于计算机内部的数据也属于网络安全的范畴。1987 年，第一个防病毒软件产品出现。随着互联网的民用化程度提高，越来越多的人开始上网，网络犯罪便开始有利可图。20 世纪 90 年代，网络安全风险暴发，早期那些炫技的个人黑客逐渐消失，黑客活动转向牟利驱动，形成了如今仍在不断壮大的网络犯罪产业。

第二阶段是 20 世纪 90 年代中后期至 21 世纪的最初 10 年，这是计算机病毒及网络黑客的暴发时期。这一时期，互联网高速发展以及社交网站、网络游戏、电子商务日益繁荣，大量计算机病毒涌现，各种网络攻击手法不断出现并呈暴发式增长，网络诈骗已具雏形。21 世纪初，在美国等互联网较早普及的国家，网络犯罪的规模、形式和恶性程度都达到了新的高度。例如，在一次网络攻击事件中，黑客获得了超过 30 亿用户的雅虎账户访问权限。2002 年 1 月，美国记录了第一起网络犯罪谋杀案，一个罪犯通过雇用黑客入侵医院的计算机系统，更改了他的竞争对手给病人开的处方，错误的处方最终导致一名病人死于一名无辜护士手中。

第三阶段的主要特征是数字化安全风险更快迭代、高智能、全覆盖。近 20 年来，网络黑色产业链日益成熟，信息泄露、网络诈骗等网络威胁迅速大范围流行，网络犯罪的规模、范围、严重程度和复杂性均呈指数级增长，且日益超越地理边界，极大地挑战了传统司法系统应对挑战的效力。2021 年 2 月初，一个黑客远程访问了美国佛罗里达州一家水处理厂的计算机系统，试图将供水中的氢氧

化钠含量增加到潜在的危险水平;2022 年 1 月,当国际油价已创近 7 年新高时,因遭到勒索软件的攻击,荷兰与比利时的几处港口无法完成石油装卸和转运,多艘油轮无法靠港;2022 年 4 月,保加利亚的国家邮政系统遭到网络攻击,柜台养老金发放业务被迫中断。

网络犯罪之所以能造成如此大的破坏,是因为它如今是一个收入颇丰的“产业”,拥有机构化的管理结构和研发预算。攻击者能够使用 AI 等高级技术工具,并能够持续加快攻击生命周期,从侦察漏洞到发起攻击,仅需几天乃至几小时。这些例子清晰地表明,数字化安全风险已经进入快迭代、高智能、全覆盖的新阶段。

在这个全新阶段,对抗安全风险已经远远不是商业机构内部的 IT 工作,而关乎每个人的幸福指数和全社会的安全稳定。2021 年 1 月,美国国务院批准设立新的网络空间安全和新兴技术局(CSET);同年 3 月,美国政府发布《国家安全战略临时指导方针》,将提升网络安全作为首要任务,并鼓励私营部门与各级政府合作,同时加大基础设施投入,保障美国免受恶意网络活动侵害。2021 年,我国通过了《国家安全战略(2021—2025 年)》,强调加快提升生物安全、网络安全、数据安全、人工智能安全等领域的治理能力,将安全提升到国家战略层面,实现国家安全战略与国民经济和社会发展规划同步。

(二)AI 时代数字安全风险的新挑战

2023 年 AI 大模型横空出世后,全球业界迅速形成共识:AI 安全风险陡增,这是全世界需要迅捷反应、共同面对的巨大挑战。目前来看 AI 安全风险主要可以分为三类:

第一是内生安全风险。从 AI 内部视角看,AI 系统和一般信息系统与应用一样,由软硬件系统组成,会存在技术本身的脆弱性、对数据的依赖性等自身缺陷导致的安全问题。比如,AI 过分依赖数据与算法,如果给数据库不断投喂带有特定价值观的数据,就会对 AI 系统产生严重的干扰,很容易产生数据偏见、观点霸权等问题,其中就涉及业界广泛关注的 AI“价值对齐”与“决策黑盒化”问题。“价值对齐”就是应确保 AI 的价值观跟人类的价值观相匹配,确保 AI 以对人类和社会有益的方式行事,不对人类的价值和权利造成干扰和伤害。目前来

看，仅靠 AI 大模型技术本身不太可能实现这个目标，需要有更多其他技术的介入。“决策黑盒化”是指难以彻底理解 AI 系统内部的结构和运转，只能通过其输出的内容来判断并得到一种外在认知。未来，如果我们真的要大规模利用大模型的能力，就必须解决黑盒问题，否则 AI 的安全永远是无根的。

第二是衍生安全风险，即 AI 系统因其自身的脆弱性被利用或被不恰当使用，从而引发其他领域的安全问题。比如，AI 系统失误可能引发一些重大安全事故。随着 ChatGPT 等 AI 大模型技术的成熟，某些别有用心的人将其作为违法犯罪活动的工具，如生成虚假新闻，深度合成伪造进行诈骗与钓鱼攻击，侵犯他人肖像权、隐私权，AI 大模型生成的内容本身也可能带有偏见。AI 衍生安全问题影响 AI 的合规使用，还涉及人身安全、隐私保护、军事与国家安全、伦理道德和法律规范等一系列与社会治理有关的挑战性问题。

第三是外生安全风险，也就是面向 AI 系统的外部入侵攻击。当前，安全风险变得更加复杂、隐蔽、强对抗和更具破坏力。一方面，数字化、网络化程度越高，可被攻击的部位、节点就越多；对数字技术的依赖程度越高，网络破坏造成的后果就越严重。另一方面，黑灰色产业链攻击模式也随着业务创新、新科技的引入，逐渐呈现专业化、多样化、智能化、产业化的特点。因而，如何将 AI 驱动的业务风控系统建设得更强更智能以应对大规模网络攻击与入侵，已成为各个企业健康发展的必要课题。

出于对 AI 大模型风险的担忧，2023 年 3 月，包括“深度学习”教父班吉奥（Y. Bengio）、特斯拉 CEO 马斯克（E. Musk）在内的多位科学家和业界顶级专家，参与签署了一封名为“Pause Giant AI Experiments：An Open Letter”的公开信。信中呼吁，因为 AI 实验室陷入了一场失控的竞赛，致力于开发和部署更强大的数字思维，但是却没人能理解、预测或可靠地控制这些大模型，所以所有 AI 实验室应该立即暂停训练比 GPT-4 更强大的 AI 系统，为期至少 6 个月。

这些问题离我们并不遥远。即使我们搁置 AI 有朝一日是否会脱离人类控制等由于“黑盒”而导致的更“哲学”的争论，在当前阶段，StableDiffusion、Midjourney 等 AIGC 工具已展现出了以假乱真的生图能力，ChatGPT 等文本生成模型也已具备高度接近人类的语言能力，一旦被别有用心的人使用，当前并没有足够有效的办法对其进行辨别与控制。随着这些 AI 技术的推广和进一步成熟，其

破坏能力将远超当下的网络欺诈等行为,给社会造成大量的混乱,给个人带来精神和财产损失。考虑到 AI 进步的速度之快、影响范围之广,AI 安全风险不可谓不紧迫。

此外,仅仅一年多时间,市面上已经涌现了各种各样的大模型。这些大模型之间如何更好地协作,共同演进而不是各行其是造成资源的浪费与社会运行的混乱,或者在协作中触发隐私保护等问题,也是未来的关键挑战。

(三)新技术与新安全风险的伴生

除了上网人数快速增加这一诱因,近 20 年网络犯罪的急剧增长,也与科技进步有关。可以说,数字化安全风险是科技发展的内在组成部分,即使互联网行业消失,所有社交媒体全部下线,安全风险依然会存在,并成为社会的潜在威胁。因为数字化并非互联网及相关行业的特有属性,而是几乎涉及所有行业、所有新兴技术。

“工业 4.0”也称“智能制造”“第四次工业革命”,即通过将信息技术集成到运营技术中,来提高工业生产的表现。这些信息技术就是我们耳熟能详的 5G、云计算、人工智能、XR、数据科学等。从全球来看,有两大因素驱动制造业的数字化转型:一是越来越多的制造型企业希望跳过中间商、服务商、代理商等,与用户直接交互;二是制造型企业希望从销售产品转为销售服务,譬如车企。

新技术工具改变了工业的运作方式,同时也导致了安全风险的蔓延,包括网络被入侵攻击,导致服务瘫痪、关键信息和数据被窃取、大范围的人身伤害。

以供应链安全为例,供应链安全风险分为两种,一种是网络遭到破坏,数据中心停止工作,可能导致整个供应链中断或瘫痪。如果遭到中断或瘫痪的是一些涉及国计民生的重要物资供应链,就会造成很大的社会影响。更危险的是,可能会发生数据篡改,导致供应链出错,譬如将不正确的药品发给了某一人群,最终造成人身伤害。

另一种供应链安全风险在安全行业更常提到,即产品开发的供应链安全风险。如今,无论开发一种硬件还是软件,作为新一代的信息化产品,它必定用到了大量的第三方开源代码。一旦这些开源产品存在安全漏洞,软件供应链就可能受到攻击,造成巨大的损失。2021 年 12 月 10 日,Apache Java 模块 Log4j 库

第一个远程代码执行漏洞被公开,在全世界都导致了广泛的攻击。①

根据 2022 年 12 月 Make UK 的一份报告,42%的英国制造商在过去 12 个月内成为网络犯罪的受害者,其中四分之一遭受了 5 万—25 万英镑的损失。根据 Security Intelligence 的数据,制造公司也成为新近流行的勒索软件最大的攻击目标,占攻击总量的四分之一,居于专业服务(17%)和政府组织(13%)之前。

当然,针对这些风险,制造业也有不少安全科技工具可以使用。比如,区块链技术可以提供分布式和不可篡改的数据存储,可用于确保供应链的透明性、防止数据篡改和验证设备的身份和可信性;AI 技术可以基于历史数据和模式来预测与防范潜在的安全漏洞和攻击,从而快速发现异常和威胁;物联网安全解决方案可以提供设备身份验证、数据加密、访问控制和漏洞管理等服务,以保护连接的设备和传输的数据;边缘计算技术能将数据处理和分析推向接近数据源的设备,从而降低数据被攻击和泄露的风险。

“工业 4.0”也并不限于制造业,事实上医疗、交通等与民生深度相关的行业,在智能化道路上面临着同样的问题。

以医疗行业为例,作为数字化程度相对较高的行业,医疗行业已经采用了大量数字技术,比如:AI 被应用于医疗图像分析、诊断辅助、疾病预测和药物研发;大数据分析被用于处理大量的患者数据,发现潜在的病例模式和治疗趋势;传感器和物联网在监测患者的健康状况、实现远程健康监护和个性化医疗服务中日渐普及;云计算帮助医疗机构存储和管理大量的医疗数据,简化数据管理流程,并提供远程访问和协作功能。

与此同时,根据中国医院协会信息管理专业委员会(CHIMA)发布的《2018—2019 年度中国医院信息化调查报告》,参与调查的 839 家医院中,仅有 43.95%通过了等级保护测评,其中三级以下医院中 75%未开展过等级保护测评。《2019 健康医疗行业网络安全观测报告》数据显示,在对 15 339 家医疗行业相关单位进行观测后,共计发现 1029 家网络服务器存在僵尸、木马或蠕虫等恶意程序的单位,6446 家单位的应用服务端口暴露在公共互联网中,4546 家单位

① 本文关于供应链安全风险的论述,来自科技部“云计算+大数据”国家重点专项课题负责人、数秦科技执行总裁兼首席科学家崔伟在接受本报告前期调研时提供的宝贵意见。

的网站存在被篡改的安全隐患,其中 261 家单位已发生网站被篡改情况。

这种状况意味着什么?2019 年,德国一家专注漏洞分析和管理的公司发现,一些服务器含有大量医疗放射图像,但它们被暴露在了公共互联网上,其中有 14 个服务器系统与中国有关,总计包含近 28 万条医疗数据。这些数据详细记录了患者的个人信息和医疗状况,攻击者借此在暗网上进行交易,获得了巨额利润。

世界已经来到了数字社会的新阶段。许多人认为数字技术是一种特色,有的行业有,有的行业没有。这种想法在 10 年前甚至 5 年前都可能是事实,当时数字技术和其他所有技术之间的界限更明显;但今天,每个行业都在走向数字化,从而提高了技术水平。交通、医疗、制造、建筑、发电等行业,在使用和依赖数字化技术方面相互联系在了一起。

为什么新兴技术要采用数字技术?或者说,为什么数字化是科技发展的内在逻辑?现代科学源于 17 世纪的科学革命,这场革命的核心特征就是将自然科学数学化。唯有如此,我们才能更精确地进行观测、计算、实验与控制。因此,数字化技术也就成了各门科学的核心工具。

以合成生物学为例,它通常是指对现有蛋白质或生物材料进行重新编程,以实现新功能。云计算和 AI 技术的兴起大大推动了这一学科的进展,生物学家可以借此处理更大的数据集,以更快的速度和更大的规模进行基因测序,从而精密地操纵和重新设计这些细胞,来驱动特定的结果、研发新的生物能源。

既然数字技术无所不在,那么数字安全风险自然也就无所不在。只要有网络、有数字信息的地方,就要有安全科技。小到微信登录,大到物流供应链系统的防火墙、疫苗生产系统的设计,都离不开安全科技。只要不出事故,我们常常感知不到这些科技的存在,但它们已经遍布生产生活的方方面面。

因此我们提出,未来有两项通用技术,一是人工智能技术,二是安全科技。技术是把双刃剑,在它以更快的速度变得更加强大时,我们不可能扔掉利刃,放弃科技——唯有以另一种科技来打造足够安全的剑鞘。

三、安全科技的关键技术能力及双重价值

（一）安全科技的内涵及关键技术能力

什么是安全科技？安全科技包括一系列旨在保护信息、网络和计算机系统免受未经授权的访问、攻击和威胁的工具、技术和系统，其市场包括硬件、软件和服务，随着 IT 基础架构向云端的演进，软件市场的占比在不断提升。

在相当长的一段时间里，网络安全（network security）即安全科技。过去 20 多年的时间里，网络安全得到了极大的发展。然而，随着数字化的深入，越来越多的细分领域涌现出来，传统的“网络安全”概念难以进行全面覆盖，而“安全科技”这个词在这样一个转型期被提出来，用以关涉多个领域的技术和理念。它的边界是不稳定的，因为新的安全风险层出不穷，安全技术一直在更新升级。

本质上，安全科技是一种伴生技术。它永远在面向新科技，面向新发展。不管哪种新技术，在推进过程中都会产生新的安全问题。新技术的发展有时候非常快，所以安全技术的发展和创新也同样在高速变动。[①]

从定义上讲，安全科技不仅超出了 network security 的范畴，也超出了国外常用的 cyber security 范畴。因为 security 更多指向与对抗相关的安全内容，而安全科技还需要包含 safety、reliability、transparency 这几个相关但不同的概念，是一个极其复杂的学科。

安全科技包含系统安全、网络安全、数据安全、业务安全、AI 安全等范畴。但在现实中，每一个技术域、每一个业务域都有自己的安全问题，都有自己的垂直分类，每一个垂直领域都很重要，区别主要在于行业侧重点不同。例如，用于保护各种应用（桌面、Web、移动等）生命周期安全的应用安全，保护各类新型终端（IOT 设备、OT 系统）安全的终端安全，保护云基础设施以及用户基于云构建的应用与数据的云安全，还有芯片安全、身份管理、威胁检测和响应、统一威胁管

① 感谢浙江工业大学网络空间安全研究院院长宣琦接受调研时提供的宝贵意见。

理、风险与合规服务等等。

目前,全球行业共有数百家公司分布在各个垂直领域,是一个较分散的市场,且更新换代速度极快,整个安全科技产业的赛道细分远超其他行业。美国拥有 80%—90%的网络安全公司,以色列、英国、法国、德国也正在培育新的明星企业。全球领先的公司有思科、微软、黑莓、Accenture、Akamai Technologies、BAE Systems 等企业,我国的深信服、奇安信、启明星辰、安恒信息、蚂蚁集团、华为、腾讯等企业在垂直领域拥有较突出的技术优势。全球权威知识产权机构 IPR-daily 发布的《安全科技专利分析报告》显示,截至 2023 年 4 月,中国是全球安全科技发明专利最主要布局的国家,共拥有专利申请数 20 445 件,为第二名美国的 2.23 倍。科技型骨干企业成为安全科技专利的主力军,蚂蚁集团、华为、腾讯、国家电网、中兴、中国工商银行、中国移动均位居全球安全科技专利申请前十名。

目前,安全科技的市场仍由传统企业占据主流,但基于云的新一代企业正在崛起。在全新的技术形势和安全需求下,新的参与者很有可能重塑这个分散的市场。

从时间上划分,安全科技的发展大致可以分成三个阶段:

第一,本地保护阶段(20 世纪 70 年代至 21 世纪第一个十年初期)。这一时期的安全科技主要基于密码学和防火墙技术,其保护的对象是网络相关的基础设施。进入 20 世纪 70 年代后,随着个人计算机的普及,各行各业都迅速采用计算机处理各种业务。计算机网络尚未大规模普及时,其主要面临的威胁是来自非授权用户对计算资源的非法使用、对信息的修改和破坏。这时,计算机安全的主要目的是,采取措施以确保信息系统资产的保密、完整和可用,比较典型的是通过操作系统的访问控制手段来阻止非授权用户的访问。

计算机网络,尤其是互联网的出现是信息技术发展中的一个里程碑。1994 年,中国科学院高能物理研究所的 IHEPNET 与国际互联网络连通,这条带宽 64K 的国际专线完成了中国全功能 IP 连接,标志着中国迈进了互联网大门。从 20 世纪 90 年代到 21 世纪第一个十年中期,我国信息化和网络迅速发展,使得围绕计算机的安全正式进入网络时代,信息系统安全(也被称为“网络安全”)产

业实现了从 0 到 1 的发展。

这一时期安全科技的作用，主要是保护信息在存储、处理和传输过程中免受非授权的访问，防止授权用户的拒绝服务，同时检测、记录和对抗此类威胁——病毒、木马、蠕虫，由此产生了对防火墙、防病毒、PKI、VPN 等安全工具的需求。20 世纪 90 年代，NASA 的研究员发明了世界上第一个防火墙。

这一阶段的另一个标志是发布了《信息技术安全性评估通用准则》，即通常所说的通用准则（common criteria，CC），后转变为国际标准（ISO/IEC 15408），我国等同采纳此国际标准为国家标准（GB/T 18336）。

第二，网络保护阶段（21 世纪第一个十年中期至第二个十年中期）。越来越多的设备接入并融入互联网，将传统的虚拟世界与物理世界互相连接，形成了一个新的空间。互联网成为个人生活、组织机构甚至国家运行不可或缺的一部分。此时，以入侵检测、反病毒和漏洞扫描等技术为代表的网络安全是主流。

第三，智能保护阶段（21 世纪第二个十年中期至今）。大数据、AI、云计算等新技术的出现和应用，改变了安全科技的面貌。这个时期的信息安全威胁来源从个人上升到犯罪组织甚至国家力量，“信息安全保障”的概念形成并成熟。智能互联设备成为新的攻击目标，系统安全风险加大。安全科技开始向智能化、自适应、自愈合等方向发展，并从技术扩展到管理，从静态扩展到动态，通过各种安全技术和安全管理措施的综合融合，构建深度的、主动的防御体系。

随着 IT 架构的演进，安全科技同样走向了无边界。在初期和发展期，政府和企业数据基本存储在私有数据中心，员工也基本在公司内办公，几乎所有的资产都在防火墙技术的保护下；到了创新阶段，数据存储转向云端和本地的混合形式，移动化办公也开始普及，办公网络的边界变得模糊；未来，大部分数据将会上云，安全科技也将从被物理硬件塑造转向被软件定义。

经过长期的发展，安全科技领域目前已经有了大量的技术储备。正如前文提到的，安全科技覆盖面极广，不同技术侧重于不同的垂直领域，且都具有各自应用场景下的重要性。根据技术的成熟度与前沿性，我们从应用价值角度将安全科技系统分为四大板块，分别是基础安全、业务安全、AI 安全、未来安全。

其中，基础安全包含传统的物理安全、系统安全、网络安全、数据安全、身份

与访问安全、密码学等，是数字世界网络中任何企业与机构都必须具备的安全防御技术能力；基础安全包括大量已经成熟应用的技术，如密码学与加密技术、防火墙、身份认证和访问控制、公钥基础设施（PKI）、入侵检测系统（IDS）和入侵防御系统（IPS）、漏洞管理和漏洞扫描、威胁情报和情报共享等技术。

业务安全是在基础安全技术保障了基础运行后，企业根据自身业务需求，针对可能遇到的业务风险而研发出的一系列技术。业务风险主要指不法分子利用业务规则漏洞和技术手段，进行数据爬取、账户盗用、交易欺诈等。因此，风控技术是业务安全技术的核心，包括风险识别、风险预警、风险管控、风险处置。近几年，风控技术加速智能化，风控模型从传统的基于规则转变为基于大数据与 AI，一些企业已开始将图技术、隐私计算等前沿技术整合进风控模型。

AI 安全是一个新领域。过去几年，通过应用 AI 来提高安全技术的效率和成功率，已经成为技术领先企业的常态。譬如通过从已有数据中学习，AI 可以更快地识别攻击的模式和趋势，从而预测未来的攻击，并配置自动响应威胁，在更短的时间内对抗网络威胁。随着 AI 大模型的诞生，AI 技术的影响力迅速扩大，不再只是助力安全，自身也成为安全保护的对象，逐渐形成全新的 AI 安全研究领域。目前，AI 安全还处于相对基础的探索阶段，主要是具备 AI 能力的企业为保障自身 AI 技术应用而进行的安全研发，如 AI 算法模型的安全、AI 原始数据的去毒清洗等。

在 AI 之外，面向未来的安全形势，另一些前沿技术也已经出现。尽管目前它们的应用规模很小，但有着相当大的潜力来改变整个安全版图，如量子信息安全技术。

整体来讲，安全科技系统具有许多尖端和前沿科技，包括：

1. AI 和机器学习。通过从已有数据中学习，AI 可以更快地识别攻击的模式和趋势，从而预测未来的攻击，并配置自动响应威胁，在更短的时间内对抗网络威胁。

2. 区块链技术。区块链提供了去中心化、安全性强的数据存储和传输机制，可用于数据完整性验证、身份认证和分布式防御。

3. 零信任安全模型。零信任安全模型意味着“从来不信任，始终在验证”

(never trust, always verify),它假设防火墙两侧的所有内容都可能有害,无论它源自何处,都要求对所有用户和设备进行严格的身份验证和授权,并实时监测和评估其访问行为。

4. 隐私计算。隐私计算是一个包含安全多方计算、同态加密、差分隐私、零知识证明、联邦学习以及可信执行环境等主流技术子项的相关技术体系,可以让数据以“可用不可见”的方式进行安全流通。

5. 物联网安全。随着物联网设备的快速增加,物联网安全成为关注的焦点,包括设备认证、数据加密、远程管理和漏洞修复等方面的技术。

6. 云安全。随着云计算的普及,云安全技术变得至关重要,包括数据加密、访问控制、虚拟化安全和云供应链安全等方面的技术。

7. 生物特征识别和多因素身份认证。生物特征识别技术(如指纹识别、虹膜扫描和声纹识别)以及多因素身份认证技术(如使用密码、指纹和面部识别),可以提供更强大的身份验证和访问控制机制。

8. 量子安全。量子安全是在未来的量子计算时代,传统加密算法受量子计算攻击而失效后,保护信息安全的一种技术。常见的量子安全技术包括量子密钥分发、量子随机数生成、量子认证、量子安全通信协议等。量子计算技术极大地提升了算力上限,并动摇了密码学的基础,目前在全球范围内引起了创新热潮,多个国家和众多行业已经投入对量子计算和量子安全的应用探索中。

(二)安全科技发展的五大趋势:智能化、零化、匿名化、弹性化和量子化

找到安全科技的共性并不难,大多数情况下安全科技发展的趋势是由最新的漏洞驱动的。因此,安全科技的最大特点就是攻防。网络犯罪者无休止地挖掘和利用新漏洞来发起攻击,安全战线一方则尝试消极或积极的防御。

有体育竞技经验的人都知道,一味防守不仅效率低下,而且表现常常难以令人满意。为了提高防御效果,需要预先对攻击者做出判断,比如:未来主要的数字安全风险会出现在哪里?哪些技术有可能成为关键?

近几年,面对新的安全威胁和新的应用需求,涌现出一大批新的安全技术,信息安全专家、中国科学院院士冯登国将这些新技术的特征概括为智能化、零

化、匿名化、弹性化、量子化。这五个特征，也可以概括为安全科技系统的整体发展趋势。

1. 智能化特征。在大模型潮流下，AI 安全必将成为主流的安全关切，该领域主要包括 AI 模型自身安全、应用导致的安全风险，以及 AI 在安全领域的应用等方面。

AI 在安全领域的应用包括防御和攻击两个方面。在防御方面，AI 赋能防御技术提升防御的能力和水平：AI 可有效提高威胁检测与响应能力；AI 可提供较高的预防率和较低的误报率；AI 可准确、快速地预防、检测和阻止网络威胁，识别并分析未知文件；AI 可克服人性的弱点，抵御以人为突破口的社会工程学攻击。在攻击方面，AI 赋能攻击技术提升攻击的精准性、效率和成功率：深度学习赋能恶意代码生成可提升其免杀和生存能力，攻击者利用深度学习模型可提升识别和打击攻击目标的精准性；AI 赋能僵尸网络攻击可提升其规模化和自主化能力；AI 赋能漏洞挖掘过程可提升其自动化水平；AI 可实现智能化和自动化的网络渗透能力。

2. 零化特征。由于云计算和虚拟化等技术的发展，网络边界越来越模糊，零信任安全架构就是基于这样的现状提出的。零信任（zero trust）不是一种技术，而是一整个框架和思想，核心是“从来不信任，始终在验证”。它的工作原理是假设每一个链接都是威胁，无论是来自外部还是内部。零安全技术主要包括零信任架构、零知识证明、零中心技术、零存在模型、零密钥协议等。

3. 匿名化特征。匿名化针对的是未来将越发凸显的数据安全问题，核心是隐私保护技术体系。部分隐私保护技术可用于解决使用中的数据安全问题，这类技术也被称为“数据使用安全技术”，如机密计算（confidential computing）、联邦学习、同态加密、安全多方计算。其中，机密计算是当前最热门的数据使用安全技术，它可提供硬件级的系统隔离，保障数据安全，为多方信息流通过程中数据的“可用不可见”提供安全解决方案。

4. 弹性化特征。弹性的目的是使系统具有预防、抵御网络攻击的能力，以及在遭受网络攻击等不利情况时仍能继续运行的能力，因此弹性是贯穿基础安全、业务安全等领域的概念。弹性安全技术主要包括弹性公钥基础设施

（PKI）、定制可信空间（TTS）、移动目标防御（MTD）、棘轮安全机制、沙箱隔离、拟态防御（MD）、可信计算等，其中弹性 PKI 被认为是新一代数字认证的基础设施。

5. 量子化特征。安全技术通常与计算能力有关，量子计算技术可使计算能力大幅度提升，从而解决现实世界中的复杂计算问题，量子计算技术将直接动摇现有安全技术的基础，因此抵抗量子计算攻击的安全设计理论、安全分析评估方法、安全解决方案及新型困难问题的寻找和优化实现等都成为当前的研究热点。

一方面，如果产学研各界能够把握这些新特征，尽快进行面向未来的安全科技研发，将有效提高数字社会的安全水平；另一方面，我们需要时刻记住，数字安全风险是和数字化转型紧紧相连的，掌握了最先进安全科技的组织和国家也将在产业发展上获得先机，从而为经济发展提供动力。

（三）安全科技的双重价值在放大

总结起来，安全科技有压舱石与助燃剂双重价值：

第一，守住技术的安全底线，防御外部风险隐患，让技术“难作恶”。例如，利用安全科技打击在线洗钱、防御 DDos 攻击、拦截非法访问等。这是数字社会稳步运行的压舱石，也是大部分人对安全科技的既有理解和定位。

第二，提高技术的安全上限，降低技术运行成本，让产业“跑起来”，让更多人受益。这一作用被许多人所忽视，是导致安全科技长期被视作成本而非动力的根源。在新技术的前期发展中，常常出现因为安全隐患而无法投入实践的问题，就好像在刹车系统被发明出来前汽车行业不可能出现。可以说，安全技术是新兴技术真正使得产业起飞的必要助推器。

当前 AI 技术蓬勃发展，适逢其时地为人类有效求解复杂问题提供了强有力的新手段，但随之而来的数据隐私、技术滥用、失控等安全问题引发了全球担忧。加强对 AI 这一新兴技术的潜在风险研判和防范，确保 AI 安全、可靠、可控，已成为 AI 产业发展的核心要素。

四、未来:安全科技将成为新型公共品

(一)“慢起飞模型”:安全科技是技术向善的集中体现

尽管我们在论述安全科技的价值时,聚焦于它对新兴科技的护航、控制以及对产业发展的促进,但必须强调的是,安全科技并不是一个纯商业范围的议题。本报告希望从两个方面厘清安全科技的公共品特征:

第一,产品运作层面。防病毒软件、入侵防御系统等商业安全系统是私有产品,但威胁情报、漏洞信息共享、隐私计算、基础设施保护等技术和平台都具有明显的公共品属性。

第二,价值观层面。长期以来,安全科技属于商业和技术界的话题,关于它的讨论一直与企业保护、IT 基础设施建设等词汇联系在一起,个人和社会在其中缺乏足够的表达。但我们认为,安全科技的完整内涵不只是对攻击者的防御体系,还包括保护平等发展、社会团结与个人尊严的技术与服务。

厘清安全科技背后的价值观,有助于我们加强对安全科技乃至整个科技价值的理解和发展路径的思考。普通人为何要关注安全科技?因为安全科技事关每个人的尊严、人际交往、健康及社会认知。

举例而言,当前相当普遍的电信诈骗,其中的个人信息从何而来?可能是犯罪分子通过非法手段购买得来的,但肯定有很大比例来自受到网络攻击的企业内部数据。尽管企业或许没有从攻击中受到直接的损害,但每个个人都承担了相应的风险。电信诈骗的主要受害人群不仅遭遇了经济上的打击,在心理上往往也承受着压力,部分人群如老年人会因此而被子女劝阻使用数字化工具,这对于人格尊严无疑是一种伤害。更进一步,即使一个人没有受到电信诈骗,他依然可能因为数据泄露而遭受网络暴力,在心理上遭受严重损害。这种心理上、人格上的伤害相当普遍却未被统计,未被纳入安全科技的充分讨论中。

相关学术研究发现,不同人群在被网络欺诈或被在线欺凌后,选择退出使用数字化设备并导致健康状况恶化的比例不同,最容易受到影响的是儿童、青少

年、残疾人、老年人等相对边缘化的群体。这说明，不安全的数字环境将扩大群体间的不平等。

平等发展角度的另一个例子是，大公司往往有更强的安全科技能力，中小型公司也认为部署安全科技能力是顶级规模的企业需要关心的议题。但现实是，勒索软件等恶意软件对中小型企业和中型市场公司会构成生存威胁，而对大型企业通常不会，因为较小的组织在遭遇打击后更可能面临重大的、公开的服务中断。例如，2019 年 5 月美国得克萨斯州一家中型钢结构制造商遭受了勒索软件的攻击，其工具和财务会计软件被永久加密，在难以支付赎金的情况下被迫破产。此外，一旦公司不得不公开处理安全问题，客户对公司的信任可能很难恢复。根据麦肯锡关于数字信任重要性的研究，近 10%的受访者表示在得知数据泄露后停止了与供应商的业务往来。

新技术会造成新的社会问题，如假新闻问题。2022 年以来的 AIGC 热潮加大了该问题的严重性。为了保护信息的真实与完整性，各国政府正在制定新的法律法规，发展新的标签系统和水印技术等，帮助人们识别虚假内容。可以预料，在 deepfake 技术的蔓延下，假新闻将比算法导致的信息茧房更进一步造成社会撕裂，加大社会成员间的沟通成本。

技术不仅是工具的集合，也反映了我们的价值观，安全科技是其中的典型。当我们讨论投资安全技术时，我们讨论的不仅是代码与数字，也是科技的伦理、社会的结构、人群的互动和个人的安全感。

为此，我们认为数字化转型应该采取“慢起飞模型”，在鼓励新技术发展、产业升级的同时，不仅要追求转型的速度，更要在社会需求方面投入更多力量，通过更好的安全科技和制度构建，尽量减少个人与社会为数字化转型所付出的代价，让起飞能惠及更多人。①

“慢起飞模型”与技术的创新冲动、商业的盈利需求之间一定程度上存在矛盾，要想切实实现“慢起飞”，需要多方的努力。一方面，要求企业真正以 ESG 的思维进行运作，不仅要投入更多资金和人力加强安全科技能力，并且要转变理

① 感谢中国社会科学院大学哲学院教授、中国社会科学院科学技术和社会研究中心主任段伟文在接受本报告前期调研时提供的宝贵建议。

念,不再仅将安全视作企业运行的成本,而是当作企业的社会责任和社会影响力的核心落脚点,是技术向善的重点领域;另一方面,这也要求安全科技真正成为公共品,由国家、社会和产业界共同推动新安全技术的低门槛、规模化应用,建立数字化安全风险公共治理空间。

(二)政企合作是安全科技成为公共品的核心驱动力

那么,如何让安全科技成为公共品?我们认为,推动力可能有四种:合规激励、商业保险、标准建设、安全思维。

一是合规激励。数据爆炸正催生越来越多技术、数据链路方面的复杂性难题,然而大多数企业往往更倾向于把资源投入 IT 研发中而不是安全保障上。目前,国内大多数企业和机构在 100 名 IT 研发人员中,配备的安全工程师不到 0.5 名。在大部分企业自身没有直接动力来投入安全防御时,政府则需要推动相关的保障投入。在这方面,欧盟 2016 年通过的《通用数据保护条例》,对于产业界起到了重大的指导意义。我国近些年在数字化安全领域同样颁布了《网络安全法》《密码法》《数据安全法》等法律,为企业建立起了基本的安全框架和安全心态。但其中多为定性规定,数字安全事件惩罚的定量及赔偿标准依然难以确定。下一阶段,合规的细化和惩罚制度的逐步实践将成为安全风险改善的重大推力。

二是商业保险。商业领域的创新是落实合规的重要助手,商业网络安全保险也能有效帮助建立数字安全事件评估及赔偿标准。网络安全保险由私人保险公司向企业和其他实体提供,保护第一方(投保人)和第三方(投保人客户)免受网络事故带来的损失。网络安全保险有两种形态,一是只覆盖网络风险的独立保单,二是作为一般保险(覆盖多种类型风险)的一揽子保险范围的一部分(如一般商业责任保险)。随着该险种的发展,市场接受率也在不断提升。

能够为合规买单的企业基本只有大型企业,从商业运作角度来说则是保险公司。由于保险公司运作要求对风险有精确的评估、监控和响应,因此商业网络安全保险市场的发展将显著提高这一领域的透明度,将安全风险从企业内部问题变成更公开的事务。另外,不同企业的保费评估能有效补充法律的定性规范,在不作为和直接触犯法律之间设置精细的处罚梯度,引导企业与机构改善自身

的安全表现。因此，商业网络安全保险有利于政府与企业、市场合作，共同攻克业态尚未成熟时期的发展瓶颈。

三是标准建设。标准建设同样是一个生态合作范畴，借助政府、头部机构与企业的带头作用，为全社会的安全技术建设提供标准。标准建设分为两个部分。第一，具体产业中的标准建设，如上文提到的 IIFAA 联盟对车联网数字钥匙的技术规范等。没有清晰的标准就会导致责任认定不清，特别是在新技术落地方面，大大增加了参与方的顾虑。第二，更高层面的安全框架规范，如美国国家标准与技术研究所制定的网络安全框架，以美国政府运行多年的《联邦信息安全管理法》经验集中输出为标准，旨在“确定现有的自愿共识标准和行业最佳实践，以将其构建成网络安全框架”，以最佳实践方式推动安全。由于参与制定安全框架标准的专家通常是中小公司无法雇用的顶级专家，因此充分利用好安全框架准则可以有效帮助中小型组织实施高质量的安全措施。

四是安全思维。我们将重点讨论两种思维转变的需求：以人为本的安全设计、运维与建设并重的安全构建。根据世界经济论坛的《2022 年全球风险报告》，95%的网络安全漏洞是由人为失误造成的。显然，个人在对抗网络风险方面承担了不相匹配的责任。一个企业乃至一个社会基础设施的网络安全不应依赖个人的持续警惕，必须要求有更多专业资源的机构通过更有效的防护机制建设来减少人为失误的风险。以人为本的安全设计就是其中一个必要的机制，它要求在整个安全风险的控制管理流程中，优先考虑参与者的体验，而不仅仅是考虑技术。运维与建设并重的安全构建，针对的是当前我国在安全方面的一个普遍倾向，即企业和政府部门构建信息化系统时，普遍存在“重建设轻运营”现象。这意味着企业在进行信息系统建设和开发时，往往只关注硬件设施建设，而忽略了持续更新维护系统的重要性。然而，数字化的特点之一就是从量变转向质变的速度很快。信息化系统运维必须有足够的运维量，这要求我们的安全构建思维从传统的硬件模式转向更偏重软件的服务模式。

（三）结语

正如数字化安全风险以及安全科技的内涵所体现出的生态融合特征，合规

激励政策的细化、商业保险市场的发展、行业标准的建设、安全思维的更新，都要求政府、企业、学术界以及公众各个层面的合作。安全科技作为一种伴生技术，本身便具有浓厚的伦理取向，需要社会各界的积极关注与参与。这种关注与参与，是我们在数字社会中安全、平等、团结与有尊严生活的必修课。安全科技作为未来的通用技术之一，尽早对它进行投资，支持我国企业在相关领域的研发和应用，将有助于掌握新兴技术发展与产业迭代的主动权，使安全科技成为建设世界科技强国的基石之一。

问题触发的算法模型响应机制探索

许正军　袁　岳*

摘要：随着数据成为各行业发展的基础性要素和战略性资源，针对问题的分析研判正由经验判断向数据分析转变，由此产生了通过数据分析提出问题解决对策的数据赋能的主要范式。本文以问题触发的算法模型为响应机制，以扎根理论的研究方法为思路，以原发问题为质性研究材料，以算法库为智能中枢，以实现从问题浮现到问题识别、问题提炼和问题响应全过程的自动化和智能化，为公共服务和商业智能领域真正实现以问题为导向发挥数据的基础要素和战略资源作用提供了一种新的思路探索。

关键词：扎根理论　数据赋能　问题　算法模型

随着数字技术的飞速发展及其在各行业的应用与渗透，数据正在不断产生和汇聚，已成为各行业发展的基础性要素和战略性资源。以问题为导向，“用数据说话，用数据决策，用数据管理，用数据创新”已成为各行业数字化转型和高质量发展的主要抓手。针对问题的分析研判也正由经验判断向数据分析转变，数据赋能正在浸润各行各业。

由此，从数据到问题，即通过数据分析提出问题解决对策，便成为数据赋能的主要范式。在此范式下，人们往往在针对场景问题的数据采集、数据处理和分析、可视化监测、趋势判断和预警预测等工程技术方面倾注大量的投入，而对问题的本原，即对问题产生的根源、内容、逻辑、性质和类别以及如何捕捉、识别、提炼和响应等方面往往缺少足够的关注。也就是说，在数据赋能应用过程中，缺乏对问题浮现、问题识别、问题提炼和问题响应等业务环节足够的渗透与智能投射，导致以问题为导向的数据赋能在上述业务环节存在“梯度消失”。

以问题为导向的数据赋能应贯穿从问题浮现到问题解决的全过程，包括问

* 许正军，零点有数。袁岳，零点有数。

题浮现、问题识别、问题提炼和问题响应(包括响应问题、确立解决方案及后续解决执行)等各环节。亦即,在数据赋能应用过程中,应从上述全链条角度,以问题数据为材料,以算法模型为智能引擎,实现数据智能在各环节的渗透与投射,消除数据赋能各环节的“梯度消失”。其中,算法模型及其响应机制至为关键。

为此,本文以扎根理论(grounded theory)基于原发问题提炼阐发的社会学研究方法,提出问题触发的算法模型响应机制构建思路,旨在通过算法模型的构建、应用和迭代,建立从问题浮现到问题识别、问题提炼和问题响应的全过程数智化响应机制。

一、总体思路

众所周知,数据承载着信息,信息蕴藏着知识,知识可以赋能智慧应用以创造价值。随着数字技术的飞速发展及其在各行各业的深入应用,各类场景问题可通过数字化技术进行数字化记录,形成问题数据。对这些问题数据进行收集(形成问题库)、处理和分析,挖掘其中的信息和知识,为问题研判、问题响应和问题解决提供决策依据,构成了以问题为导向的数据赋能的基本逻辑。

但在实际过程中,上述大多只停留在数据采集、数据处理和分析、可视化监测、趋势判断和预警预测等工程技术方面(图 1),对于问题的本原,如问题产生的根源、内容、逻辑、性质和类别以及如何捕捉、识别、提炼和响应等方面缺乏足够的考虑,即在数据业务化层面做得不够充分。

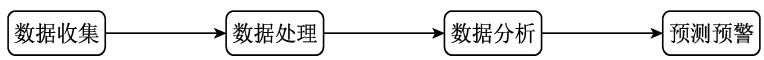


图 1 从数据采集到预测预警的示意图

问题触发的算法模型响应机制将立足于从原发问题的本原探究出发,从问题浮现、问题识别、问题提炼和问题响应等各环节着手,以问题数据为材料,以算法模型为智能引擎,建立从问题浮现到问题识别、问题提炼和问题响应的全过程数智化响应机制(图 2),实现“数据赋能”贯穿从问题浮现到问题解决的全

过程。

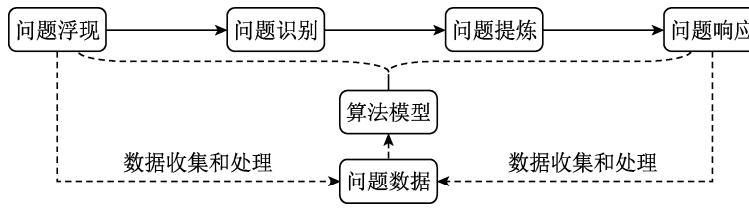


图 2　问题触发的算法模型响应机制总体思路

从图 2 可见，问题浮现是整个机制的起点，问题响应是整个机制的终点，二者都是问题数据的来源——前者是问题的初始来源，后者是问题解决成效的反馈，并且构成了一个问题从浮现到解决的闭环。通过对问题浮现和问题响应的数据收集和处理，形成问题数据库，就形成了问题触发的算法模型响应机制构建之源，这也是问题“业务数据化”的过程。问题识别和问题提炼是问题“业务数据化”的核心，是从问题浮现到问题响应的智慧转换器，而智慧之源来自算法模型。算法模型封装了多种基础算法和针对各类垂直应用、特定任务解决能力的智能引擎，这些算法和引擎通过对问题数据的分析和挖掘，高效、精准地识别问题的来源、特征、内容、类别和其他相关属性（如紧急事件、突发事件、热点事件或其他难点痛点堵点事件等），并根据问题识别的结果辨析出那些符合实际的可触及、可控制、可解决的问题，同时提炼出问题中的相关概念和命题，结合历史经验（对策库）和一些标杆做法，提炼出相应的问题解决理论体系和响应对策，从而实现“数据赋能”贯穿从问题浮现到问题解决的全过程。

二、问题浮现

问题浮现是整个机制形成的起点，也是初始问题数据收集的来源。无疑，了解问题浮现的起因与渠道至为关键。

不论是公共服务领域，还是商业领域，问题浮现的起因大多来自需求侧的各类诉求，包括投诉、表扬、建议、咨询、求助与评价等，也有来自第三方的调查或对供给侧的服务监督过程中的主动发现。问题浮现的渠道也呈现多源性特点。

以公共服务为例，问题的浮现有来自企业和群众的主动反馈，如通过 12345 政务服务便民热线，或通过政务服务好差评平台，或通过线下政务大厅专设窗口（如“不办事”窗口）；也有来自大调研或第三方调查过程中的主动发现，如通过结构化或半结构化问卷的发放与问题采集；还有来自政府部门在重点事件关注或执法过程中的发现，或是如“双随机、一公开”执法监督过程中的主动发现；也包括对相关互联网办事平台与其他互联网媒体监测过程中的问题捕获；等等。问题的内容可能涉及方方面面，如政务服务、营商环境、市场监管、生态环保、公共管理、公共安全、公共服务等。问题的焦点可能涉及一些长期以来的难点、堵点和痛点，也可能涉及一些突发、并发或热点事件，抑或涉及政府部门某些重点关注的事件，等等。

不同于以往的仅从各业务系统收集数据并试图从数据中挖掘问题，深入了解与分析将有助于多渠道、多角度捕获问题的浮现，特别是对于潜在的隐性问题的发现，有助于拓展问题数据的收集渠道（如大调研、第三方调查、访谈、会议协商、检测记录等），提升数据收集的质量。同时，对问题浮现的深入了解将增强对问题浮现机理的全面掌握和深入了解，为后续的问题识别、问题提炼和问题响应创造良好条件。事实上，问题浮现的渠道、主题、内容和主体之间的矩阵图谱构成了具象化的问题浮现的机理特征（图 3）。

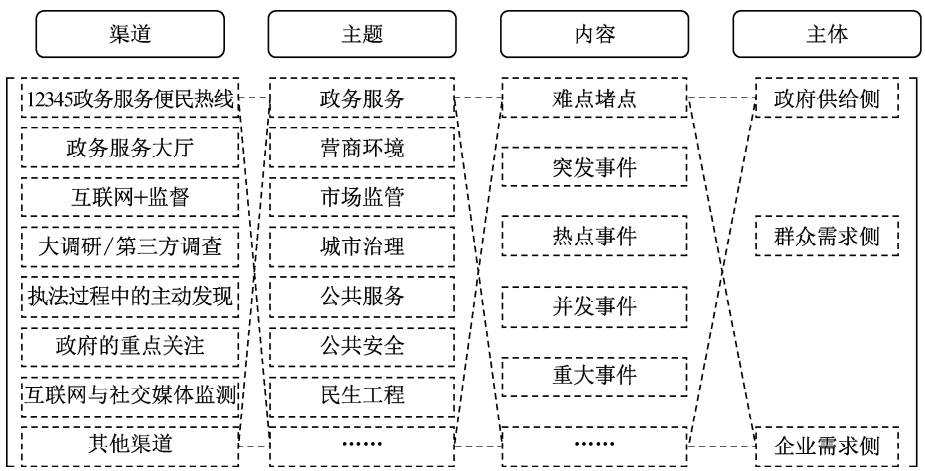


图 3　问题的渠道、主题、内容和主体之间的矩阵图谱

三、问题识别

问题识别在于挖掘问题的特征和识别问题的类型，这将采用扎根理论质性研究方法来实现。首先，对问题数据进行详细的分类和标签化处理，识别问题的基本概念和特征（简称“初始编码”）；其次，将这些概念和特征进行关联，分析它们之间的逻辑联系、条件关系和因果机制，找出它们的主题范畴（简称“主轴编码”）；最后，确定一个核心主题，围绕这个主题整合主题范畴和概念，构建问题的核心结构（简称“选择性编码”）。上述编码可挖掘问题的特征和识别问题的类型。

以公共服务领域中的营商环境问题为例。如前所述，营商环境问题来源众多：有来自政务数据资源库，也有来自企业访谈、营商小组讨论、相关文件资料等渠道，还有来自 12345 政务服务便民热线、政务好差评、投资服务中心、招商过程中的观察记录，以及其他政务服务业务系统等。基于对营商环境问题浮现机理的理解进行多源数据收集和处理，形成营商环境问题数据库，以此作为问题识别的质性研究材料。

首先，寻找并标记营商环境问题数据中供需双方关于营商环境提升和优化的初始概念与特征内容。例如，从企业需求侧可能会发现如“希望更加公平的市场环境”“司法程序复杂”“项目货款拖欠严重”“招工难”等众多与营商环境需求相关的诉求概念与特征内容，从政策供给侧可能会发现如“关于进一步加强公共资源交易监管的指导意见”“知识产权保护”“信用监管”“人才培训与人才引进政策”等众多与营商环境提升相关的诉求概念与特征内容。

其次，对这些初始概念或特征内容进行分类和关联，归并或提炼到每一个主题范畴中。例如，将“希望更加公平的市场环境”和“关于进一步加强公共资源交易监管的指导意见”归并到“市场环境”这个主题，将“司法程序复杂”和“知识产权保护”归并到“法治环境”这个主题，将“项目货款拖欠严重”和“信用监管”归并到“信用环境”这个主题，将“招工难”和“人才培训与人才引进政策”归并到“要素环境”这个主题，等等。

最后，确定一个核心主题，整合各个主题范畴和概念。例如，确定“营商环

境优化和提升行动”这个核心主题,将“市场环境”“法治环境”“信用环境”“要素环境”及其他主题、概念进行整合,形成针对营商环境优化和提升行动的基本理论框架(图 4)。

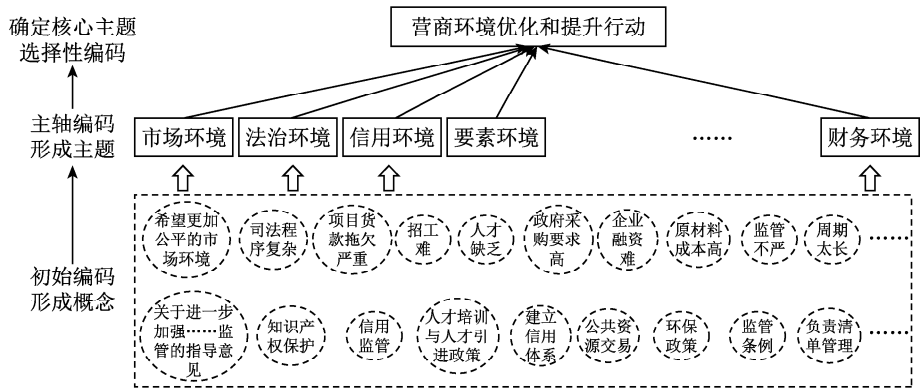


图 4　针对营商环境问题的编码与识别(1)

实际过程中,可根据需要进一步将主题范畴细分为多级子主题范畴。例如,将“市场环境”进一步细分为“市场准入”“市场监管”等,将“法治环境”进一步细分为“依法监管”“法律诉讼”等,将“信用环境”进一步细分为“社会信用”“信用监管”等(图 5)。

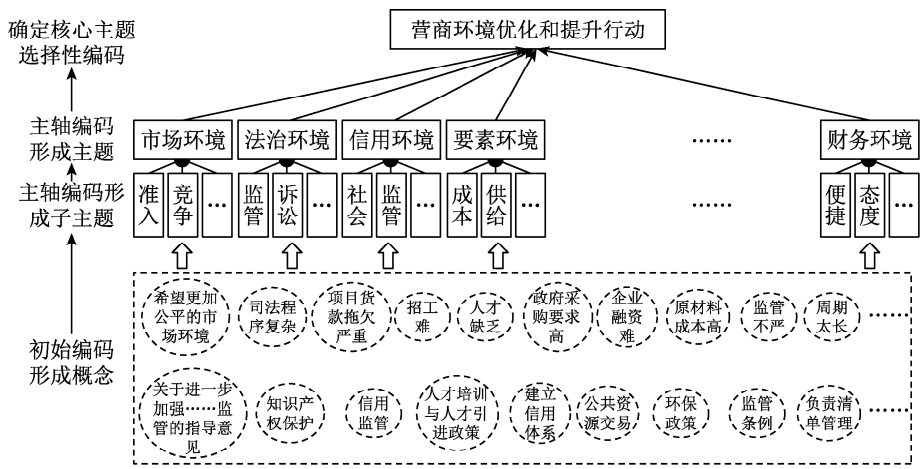


图 5　针对营商环境问题的编码与识别(2)

考虑到与监测和评价指标的描述相一致，在问题识别过程中，可借助相关监测与评价的指标名来匹配编码过程中的相关概念和主题范畴。针对营商环境问题，可参照营商环境评价指标体系来进行。

四、问题提炼

在问题识别形成概念和主题基本理论框架的基础上，进一步对这些概念进行关联和匹配，实现问题提炼，为下一步的问题响应提供建议与理论指导。问题提炼主要涉及三个方面的核心内容：提炼概念中的异常情况（形成类似分类算法中的负例数据集）；辨析异常情况的可触及性、可控性和可操作性，并结合其他特征变量分类排序形成问题序列（相当于对负例数据集进行等级分类和排序）；进一步对问题进行深度挖掘，形成新的概念和命题，构建问题新的理论范式，并结合经验对策和一些标杆做法生成问题响应的对策与建议（相当于对负例数据进行分析挖掘后形成的对策与建议）。从技术角度而言，这三个核心内容构成了问题提炼器的三大核心功能（图 6）。

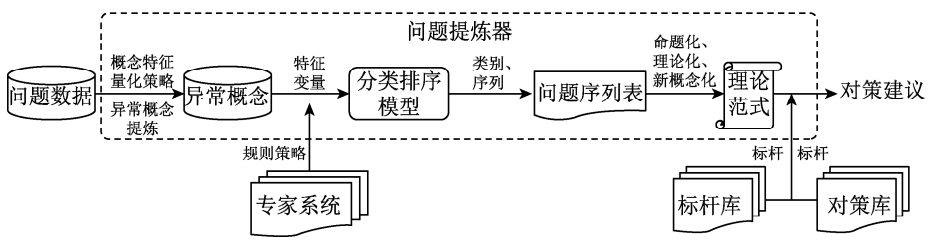


图 6　问题提炼器的三大核心功能

仍以营商环境问题为例。企业侧的需求和政府侧的供给是营商环境问题的两个方面，因而在问题识别的基础上，可对概念进行供需之间的关联和匹配，以发现供需之间哪些概念具有一致性，哪些概念存在部分差异，哪些概念存在显著冲突和矛盾（对于新出现的概念，技术上可视为此类），等等。例如，企业需求侧的“希望更加公平的市场环境”与政府供给侧的“关于进一步加强公共资源交易监管的指导意见”之间可能是一致的，“司法程序复杂”与“知识产权保护”之间可能存在部分差异，而“项目货款拖欠严重”与“信用监管”、“招工难”与“人才

培训与人才引进政策”之间可能存在显著冲突和矛盾(图 7)。

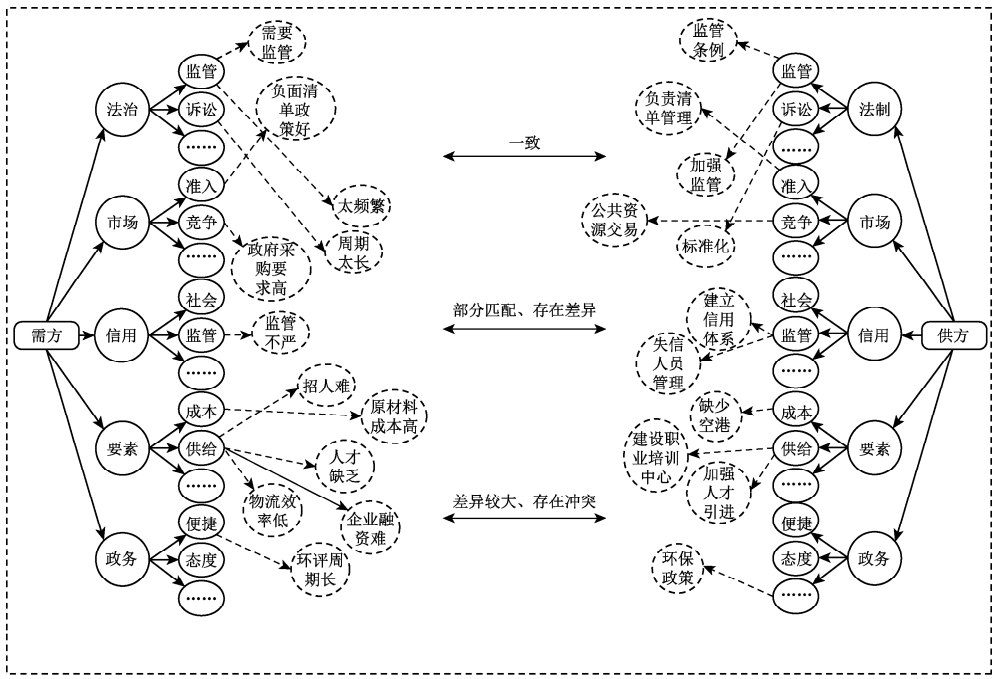


图 7　营商环境问题中的异常概念提炼示意图

其中,一致性说明营商环境的提升行动是满足需求的,存在部分差异的说明营商环境仍需要进一步完善和优化,而存在冲突和矛盾的或新出现的概念,则需要重点关注。

鉴于并非所有的异常概念都具有合理性和可解性,因而问题提炼的第二个核心内容则是进一步辨析问题的可触及性、可控性和可操作性,并根据轻重缓急与其他相关特征变量(如差异程度、紧急程度、突发/并发/热点、社会影响、合理性、可解性、难易程度等),通过分类排序模型生成问题序列。这部分工作需要借助有关规则和专家系统来进行。

问题提炼的第三个核心内容将是进一步对问题进行深度挖掘,形成新的概念和命题,构建问题新的理论范式,并结合经验对策和一些标杆做法生成问题响应的对策与建议。具体而言,通过进一步对初始问题进行深度挖掘,从初始问题的表述中提炼概念、实体、关系、属性、事件、特征等关键要素及其之间的逻辑关

系、因果机制和知识图谱，构建一系列新的命题范畴。同时，通过重新挖掘初始问题的关键逻辑和规则，进一步提炼、综合和整合不同表述的同类概念，形成一系列新的概念表述；结合前述的异常分析和问题序列，构建新的核心主题。新的概念表述、新的命题和新的核心主题组合形成了初始问题的新的理论范式。在此基础上，再结合经验对策和一些标杆做法生成问题响应的对策与建议。

五、问题响应

在问题响应环节，借助问题提炼生成的对策与建议，有针对性地制定相应的解决方案，为问题解决提供行动指南。

在制定解决方案之前，首先需要根据问题提炼阶段提炼出的有关问题要素，如问题的内容、性质和主体等，决定是否需要进一步与问题主体（问题当事人或其他相关人）进行沟通，以挖掘问题的真实原意；或对照比较是否属于领导重点关注的重大事件之列。同时，根据不同的问题内容、性质和主体，挖掘问题的核心目标是什么，需要投入哪些资源和条件，需要哪些法律法规支撑和保障措施，需要履行哪些决策程序，执行周期会有多长，执行后是否会达到问题的核心目标。将这些决策变量组合在一起进行推演和模拟，当预期效果能达到问题的核心目标时，便可形成解决方案和行动指南。对于达不到预期效果的问题，或是一时找不到好的对策的问题，可以借鉴其他地方的问题解决标杆来确定自己的解决方案。

总之，在问题响应环节，首先需要基于问题提炼的对策与建议，确定问题解决需要的要素条件，形成问题解决的时间路线图，组合不同的条件变量进行推演和模拟，形成有针对性的问题解决方案，这便构建了问题响应和问题解决的基本构架与主要逻辑。

作为闭环，对于问题响应和问题解决的成效需要进一步跟踪，形成问题响应和问题解决的跟踪库。对于成效不高的解决方案，需要进一步修正有关条件变量和时间路线图；对于达到问题核心目标的解决方案，便可形成具有参考价值的对策库，为日后类似问题的解决提供自动响应机制（图 8）。

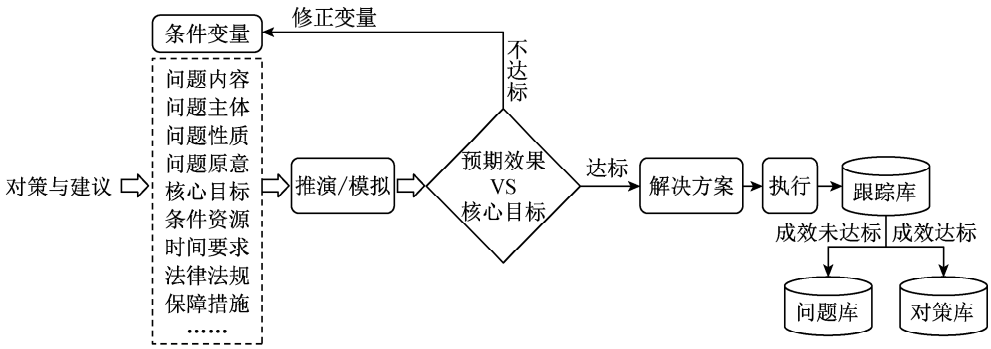


图 8　问题响应过程示意图

六、关键技术

问题触发的算法模型响应机制，仍然离不开数据收集、数据处理和数据分析等一些常用大数据技术的支持，包括数据治理、标准规范和数据安全等方面的技术应用。从前文可知，算法模型为智能引擎，是实现数据赋能贯穿从问题浮现至问题识别、问题提炼和问题响应全过程的关键。因此，本文着重讨论算法模型构建的关键技术。

通常，基于扎根理论的研究可以借助一些成熟的软件工具来实现，如 Nvivo、MAXQDA 和 QDA Miner 等。尽管这些软件在编码方面表现出较强的性能，然而在效率上仍属于“作坊式”定制，在准确度上仍需大量的人工干预。特别是在理论饱和环节，需要重复收集和分析数据，自动化和智能化程度不够。

近年来，机器学习尤其是深度学习技术得到了长足发展。特别是随着 ChatGPT 的出现，以 GPTs 为代表的生成式人工智能正在应用到各行各业中。本文讨论的算法模型关键技术，将全面采用机器学习算法来实现（图 9）。

在问题浮现环节，可以通过“超能交互”、智能语音识别技术及图像识别、图像分割和图生文大模型技术，实现对问题浮现的捕获与问题数据的生成。

在问题识别环节，可以组合运用分词技术（如 N－gram 模型、隐 Markov 模型、最大熵模型和神经网络算法）、TF－IDF 等文本特征选择与向量化模型技术

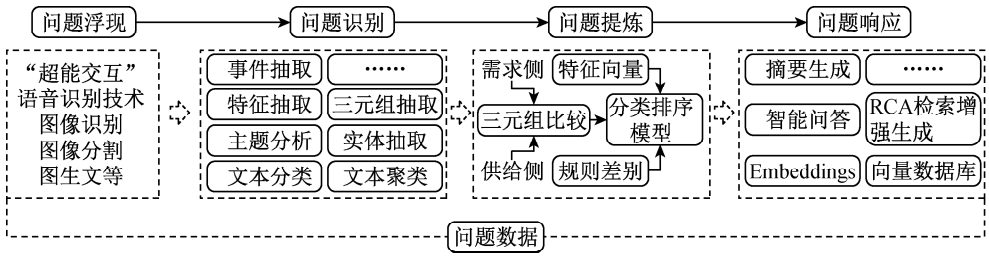


图 9　机器学习算法在问题触发的算法模型响应机制中的应用

选择问题数据中的特征和概念；运用词向量技术（如 Word2Vec）和潜在狄利克雷分布（LDA）主题分析技术，可提取问题数据中的关键词和若干主题，并可标示主题与关键词之间的权重情况，建立主题与关键词之间的关系，识别关键词之间的语义关系和相似度；通过知识智谱技术（如实体抽取、实体对齐、三元组抽取、事件抽取等技术），挖掘各种概念之间的关联。以上技术的运用，再加上文本分类、文本聚类等，为问题识别提供了丰富的技术手段。

在问题提炼环节，通过对实体对齐后的供需三元组属性值进行比较分析，以识别供需之间同一范畴内的概念是否一致，或是否存在差异，是否存在显著冲突和矛盾等情况。对于异常概念，可根据问题的可触及性、可控性和可操作性及其他相关特征变量（如差异程度、紧急程度、突发/并发/热点、社会影响、合理性、可解性、难易程度等），通过分类排序模型生成问题序列；可通过如主成分分析（PCA）或非负矩阵分解（NMF）技术进一步抽取问题数据的特征和主题，形成新的概念和命题，以形成问题的新理论范式。结合历史经验和标杆，利用类 ChatGPT 生成模型生成问题响应的建议和对策。整体上，上述技术的有机组合形成了一个问题提炼器。

在问题响应环节，可通过历史数据训练一些垂直领域大模型，建立问题解决方案中的条件变量、时间路线图与问题解决成效之间的对应关系，并以此大模型作为解决方案的扮演和模拟器，提升问题解决方案制定的科学性。其他如 Embeddings 模型、RGA 检索增强生成和向量数据库技术等，可为问题响应提供智能问答和文档摘要生成能力，助力问题响应过程中的自动化和智能化。

实际过程中,这些算法模型的集合构成了问题触发的算法模型响应机制的算法库,作为贯穿从问题浮现到问题识别、问题提炼和问题响应全过程的智能中枢(图 10)。

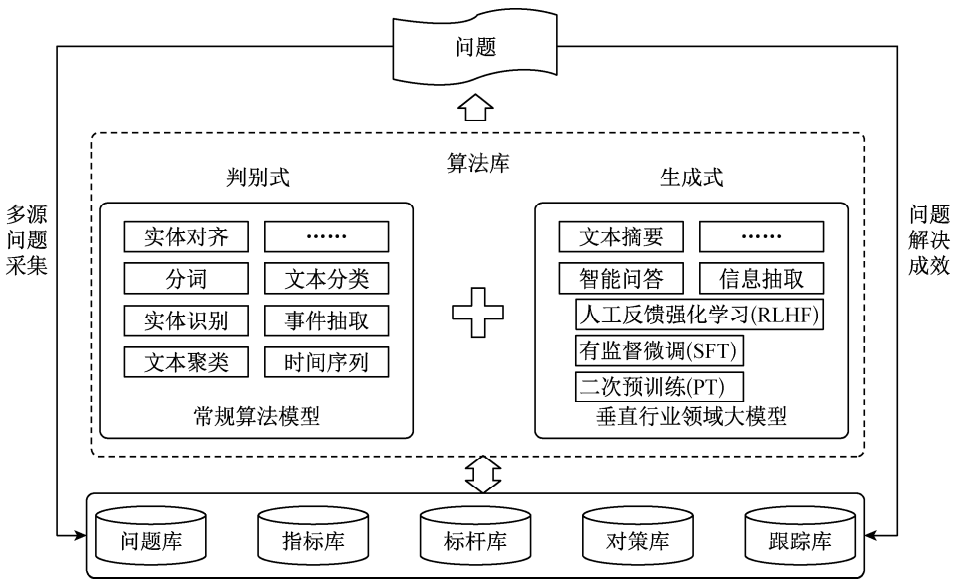


图 10 从问题浮现到问题响应的智能中枢——算法库

由此可见,问题触发的算法模型响应机制,以扎根理论的研究方法为思路,以原发问题为质性研究材料,以算法库为智能中枢,从而实现从问题浮现到问题识别、问题提炼和问题响应全过程的自动化和智能化,为公共服务和商业智能领域真正实现以问题为导向发挥数据的基础要素和战略资源作用提供了一种新的思路探索。

实践过程中,需要重点把握三个方面的工作。

第一是问题体系构建和模型准备。针对某一行业或领域,坚持以原发问题为导向,厘清问题浮现的源头,收集历史问题多源数据,通过垂直行业大模型训练并建立覆盖问题浮现监测、问题识别、问题提炼和问题响应能力的算法模型(库),并构建问题标准化体系(指标库)。以 12345 服务便民政务热线场景为例,问题浮现的源头主要有 12345 电话呼入、各级政府的微信与微博、各类政务服务平台的满意度评价和好差评、大调研、第三方调查、各类互联网媒体及其他

等。对这些渠道的历史数据进行收集，建立 12345 政务服务便民热线行业大模型，打造覆盖 12345 政务服务便民热线各类问题浮现监测、问题识别、问题提炼和问题响应能力的算法模型（库）（模型能力 APIs），并构建 12345 政务服务便民热线问题标准化体系（各类问题指标库）和对策库。

第二是问题提炼器的构建。在整个算法库的构建过程中，作为初始问题新的理论范式构建的重要引擎，问题提炼器在整个算法模型响应机制中起着至关重要的作用。问题提炼器的能力和性能决定了新概念和新命题提炼的质量，也决定了新的理论范式的先进性，从而也决定了对策与建议的科学性和准确性。

第三是针对具体细分场景问题的应用。应用所建立的问题标准化体系、对策库和算法模型（库），监测具体细分场景问题的浮现，通过问题识别、问题提炼（问题解析）和问题响应（推演和模拟），对应到具体的解决方案。问题解决方案的成效将被跟踪记录（跟踪库），作为闭环，不断优化迭代算法模型的应用能力。

数字健康的社会技术伦理*

——生命伦理学进路的批判和延伸

詹姆斯·肖　约瑟夫·多尼娅　著**

何　丽　译***

摘要:数字技术的广泛应用,在医疗保健和公共卫生领域引发了重要的伦理问题。理解这些伦理问题需要一种超越技术本身的视角,将技术所处的社会技术系统包括在内。社会技术系统是指涵盖材料设备、人际关系、组织政策、企业契约和政府法规的广泛集合,它们塑造了数字健康技术的接受和使用方式。评估数字健康技术的生命伦理学进路,通常受限于技术自身特征引发的伦理问题。仅考虑技术功能的伦理视角,不足以评估技术应用对于照护环境及其所属更大范围的健康相关生态系统所产生的广泛影响。首先,回顾了数字健康的生命伦理学研究的现有进路,并借用设计伦理与科学技术学(STS)的概念批评这些进路的狭隘观点;其次,对数字健康技术应用于医疗保健环境中所产生的社会技术系统加以论述,概述了在此广阔视角下对数字健康技术展开全面的伦理分析需要注意的各类因素;最后,从社会技术视角概述了社会正义对伦理分析的重要意义。

关键词:数字健康　生命伦理学　科学技术学(STS)　远程医疗　伦理学

一、引言

世界各地众多的政策制定者、医务工作者(health care provider)、研究人员

---

* 本文原标题为"The Sociotechnical Ethics of Digital Health: A Critique and Extension of Approaches From Bioethics",原文见 https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2021.725088/full,经授权翻译发表。

** 詹姆斯·肖(J. Shaw),多伦多大学卫生政策、管理和评估研究所(IHPME)。约瑟夫·多尼娅(J. Donia),多伦多女子学院医院。

*** 何丽,复旦大学哲学学院。

和技术用户，对于数字技术（digital technologies）促进健康和改善医疗保健（health care）的希望正持续增长（Henwood & Marent，2019；Marent & Henwood，2021）。患者与公众对于在医疗保健系统中使用数字技术通常持积极观点（Lupton，2017a；Haluza，Naszay & Stockinger et al.，2016；Montagni，Cariou & Feuillet et al.，2018）。在政策层面，数字健康（digital health）也被视为实现更有效、更便捷的医疗保健服务的战略（Greenhalgh，Procter & Wherton et al.，2012；Shaw，Jamieson & Agarwal et al.，2018）。世界卫生组织（WHO）发布了首份数字健康全球战略报告[①]，并且制定了若干指南以便指导数字健康技术在现实中的评估和应用（Auerbach，2019；WHO，2016；Murray，Hekler & Andersson et al.，2016）。尽管在医疗保健中实现对数字技术的有意义使用持续存在挑战（Shaw，Agarwal & Desveaux et al.，2018），但人们普遍乐观地认为，数字健康技术在促进健康和改善未来医疗保健方面将发挥重要而积极的作用（Bayram，Springer & Garvey et al.，2020）。

对数字健康潜力的乐观情绪以及对推进数字健康议程的承诺，只代表对数字健康本质及其影响的部分看法。新冠疫情大流行催生了一系列研究工作，揭露并使公众意识到数字技术在加剧健康不正义方面的潜在影响（Veinot，Mitchell & Ancker，2018；Latulippe，Hamel & Giroux，2017），以及大型科技公司对公共卫生政策的影响（Sharona，2020），等等。此外，数字技术（资源）的分配极大改变了人们看待和建构生活的方式，这些变化与健康和医疗保健的实践和制度密切相关（Lupton，2017a）。

这些新近的观察提出了一些关键问题，涉及数字健康的众多应用后果，以及社会所期望的数字健康的未来发展方式。这些问题属于规范性问题，与人们对数字健康技术在促进健康和提供医疗保健方面应该发挥的作用的设想有关（Lupton，2017b；Felt，2015）。然而，关于数字健康的规范性基础，以及数字健康对健康与医疗保健所施加影响的研究和著述还存在局限。本文将从生命伦理学（bioethics）视域探讨关于数字健康的现有观点，并提出一种思考这一重要议题的替代性进路。尽管在宽泛的应用健康伦理（applied health ethics）领域，如公

① https://www.who.int/westernpacific/health-topics/digital-health.

共卫生伦理(public health ethics)领域,也存在其他观点(Benatar, Upshur & Gill, 2018),但我们特别聚焦生命伦理学,因为它是对医疗保健与医学问题开展伦理分析的主要进路(Fox & Swazey, 2005)。因此,我们的批评仅针对从传统生命伦理学视角分析数字健康的研究成果,但这些批评也与健康和医学以外的应用伦理学(applied ethics)相关进路有关。事实上,生命伦理学并没有严格的界限,文中提到的许多方法同样适用于生命伦理学以外的其他应用伦理研究领域(如计算伦理学)。

数字健康是指一系列随时间推移而不断变化的技术和实践,伴随新技术的出现而不断发展。文中对"数字健康"的定义与本杰明·马伦特(B. Marent)和弗利斯·亨伍德(F. Henwood)的定义保持一致。他们将以下四种技术辅助照护形式纳入数字健康的范畴(Marent & Henwood, 2021):远程医疗(远程同步或异步照护)、电子健康(eHealth,健康信息的搜索和交换)、移动医疗(mHealth,为健康相关原因使用移动数字设备)和算法医学(将数据科学和人工智能的进展融入医疗保健)。基于此,我们对"数字健康"的定义,包括使用数字技术进行自我追踪(self-tracking)或自我照护(self-care),健康信息搜索和交换以及直接提供健康和社会照护(health and social care)。

二、数字健康的生命伦理学:一种批判

鉴于"数字健康"这一术语仍相对较新,为了更好地区分被纳入这一领域的众多技术应用,数字健康的生命伦理学分析直到最近才开始发展也就不难理解了(Brall, Schröder-Bäck & Maeckelberghe, 2019)。然而,一些现有的出版物已经尝试推进数字健康的生命伦理学研究和实践。与此同时,该领域正持续激发研究者的研究兴趣,目前已经形成研究群体,提出了各种认识与研究路径以了解生命伦理学这一研究领域(Milosevic, 2019; Nebeker, Torous & Ellis, 2019; Schmietow & Marckmann, 1029; Ruotsalainen & Blobel, 2020)。作为数字健康和创新伦理领域的研究者,我们密切关注此类研究。在本节,我们将这些研究分为三类,对其分别进行描述并给予总体性评论。

我们将第一类称为"应用伦理理论"(applying ethical theory)。在此类研究

中，学者们采用现有伦理理论的视角来评估与数字健康在规范意义上相关的问题子集（Brall，Schröder-Bäck & Maeckelberghe，2019；Winters，Venkatapuram & Geniets et al.，2020）。最常见的是原则主义（principlist）进路，依靠一系列生命伦理原则的指导来评估特定人类活动领域的伦理影响（Nebeker，Torous & Ellis，2019；Schmietow & Marckmann，2019；Joerin，Rauws & Fulmer et al.，2020；Mulvenna，Bond & Delaney et al.，2021）。生命伦理学领域是由原则主义的伦理思维方式主导的（Fox & Swazey，2005；Hedgecoe，2004），因此，数字健康的生命伦理学研究同样由原则主义主导也不足为奇了。沿着这一进路，研究者往往不使用复杂的理由来支持伦理理论的特定取向，而主要专注于将理论应用于数字健康的实质性问题。比如，莫里斯·马尔文纳（M. Mulvenna）等人曾撰文探讨在健康相关领域使用数字表型的伦理问题。在论文中，他们简明扼要地指出，“医学的四大伦理支柱是自主、有利、无害和公正，它们在数字表型的普及化过程中不应被忽视”（Mulvenna，Bond & Delaney et al.，2021），并且在伦理分析中继续将自主原则作为首要关注点。此类研究从常见的生命伦理学理论视角，阐明了数字健康中一系列已明确界定的规范性问题的各个方面。其中，最突出的包括隐私、安全、数据治理以及使用数字健康技术所产生的利益和责任的分配（Sharon，2020a；Winters，Venkatapuram & Geniets et al.，2020；Char，Abràmoff & Feudtner，2020）。

第二种类型可称为“将伦理转化为实践”（translating ethics for practice）。这一类型的研究聚焦于增强数字健康生态系统中的利益相关者以有意义的方式理解与应用生命伦理学概念的能力。将生命伦理学转化为实践，并不是从特定的伦理视角去分析伦理问题，而是将具有伦理意识的声明或原则与开发或实施数字健康技术的实践联系起来。例如，佐兰·米洛舍维奇（Z. Milosevic）详细阐释了义务论并概述了具体的义务论概念如何直接与软件设计过程相关联（Milosevic，2019）。亦有研究旨在进一步简化生命伦理学的原则主义进路，并说明其与数字健康技术设计各个方面的联系（Joerin，Rauws & Fulmer et al.，2020）。这一进路下的研究，旨在简化和说明生命伦理学理论（bioethical theory）对构建和部署数字健康技术的实际工作的影响（Joerin，Rauws & Fulmer et al.，2020）。

我们将第三种类型称为“识别伦理危害”(identifying ethical harms)。这一进路旨在识别和描述数字健康领域存在的伦理方面的相关危害或规范性问题。被识别的伦理危害在其与技术本身的接近程度上有所不同。伦理危害既包括更接近技术的问题,如数字健康技术的隐私或信任问题(Ruotsalainen & Blobel, 2020; Vayena, Haeusermann & Adjekum et al., 2018);也包括其他远离技术本身的问题,如世界各地在获得和使用数字健康技术中存在的技术资源可及性问题(Winters, Venkatapuram & Geniets et al., 2020)。尽管此类研究的重点是数字健康的危害、问题或挑战(Morley, Machado & Burr et al., 2020),但这些问题往往与能够解决危害的积极概念相联系。例如,艾菲·瓦耶娜(E. Vayena)等人指出信任问题是数字健康技术面临的一大挑战,而(与之相关的)问责制从长远来看则能促进信任在这一领域的产生与发展。此类研究使得与数字健康相关的规范性问题(如隐私和安全)备受关注,也推动了对未决问题的深入思考(Winters, Venkatapuram & Geniets et al., 2020)。

这里总结的三种进路,均为全球讨论数字健康技术所带来的伦理挑战和应对的潜在策略做出了重要贡献。具体地说,它们阐明了隐私、自主、安全、同意、透明、问责和公平的重要性,并探索了数字健康治理的丰富路径。然而,它们也受到了重要的批评。这些批评建立在生命伦理学作为一个研究和实践领域所普遍遭遇的批评之上,对我们自身理解数字健康伦理具有借鉴意义(Fox & Swazey, 2005; Fox, 2016; Hedgecoe, 2004; Hedgecoe, 2010)。我们的批评基于对生命伦理学作为健康和医学领域的应用伦理学的两点核心观察,这两点观察直接适用于数字健康伦理学的相关文献。首先,生命伦理学建立在一个基本信念之上,即存在道德普遍性(moral universals),它在本质上不受世界各地不同地区的社会、文化和政治现实的影响(Fox, 2016; Fox & Swazey, 2008; Myser, 2003)。其次,对于健康或医疗技术取得的某种进步,生命伦理学的常见做法是接受由相关临床医疗机构或技术制造商设定的边界(Hedgecoe, 2004; Hedgecoe, 2010)。我们将就上述数字健康伦理学相关文献中的观点逐一进行讨论。

前两种进路依靠现有伦理理论来解决数字健康领域的各种问题。这些研究几乎没有涉及对特定伦理进路的详细论证,当然它们也不能;一篇应用伦理学论文从根本上来讲并不涉及特定伦理理论本身在哲学上或理论上的证明。然而,

依赖传统生命伦理学进路的弊端也日益显现。社会科学对生命伦理学的批评清楚地说明了预设普遍道德的问题，正如蕾妮·福克斯（R. Fox）和朱迪思·斯瓦泽（J. Swazey）所清楚表明的那样："这一领域致力于确定和促进构成'普遍道德'（common morality）的普遍伦理原则。"汤姆·比彻姆（T. Beauchamp）和詹姆士·邱卓思（J. Childress）将其描述为："所有在道德上严肃的人……在所有地方……分享的一套规范。"（Fox & Swazey, 2010）这种伦理取向忽视了权力和文化在塑造道德信念方面的基本作用（Mattingly, 2012；Mattingly & Throop, 2018），也忽视了生命伦理学中充斥着强化现有权力体系状况的预设。我们的目标是，承认这些对生命伦理学话语的影响，促进对以往伦理研究中的种种预设，以及我们试图推进的特定规范性立场的自我批判式分析。

"识别伦理危害"受到与前两种进路相关但不同的批评：生物伦理学研究者倾向于接受特定技术支持者围绕伦理话语所设定的界限（Hedgecoe, 2010）。亚当·赫奇科（A. Hedgecoe）在总结遗传药理学领域的生命伦理学研究后，曾清楚地提出这种批评（Hedgecoe, 2010）。他指出，研究者普遍接受科学家关于药物遗传学进展在医疗保健中适当作用的主张。他说："很明显，这些研究者可以对这些所谓的'科学说法'提出质疑。只是他们并没有这么做。（我们）也不清楚为什么这些研究者允许他们的话语停留在当前的范围之内，并且如此不愿以新的方式思考药物遗传学提出的伦理问题。"（Hedgecoe, 2010）

在数字健康的生命伦理学方面的工作，似乎在很大程度上受到同样的批评。与数字健康技术相关的一系列常见问题（一如前文所总结的）经常在伦理角度被定义和讨论，但这些伦理危害以外的其他问题却极少遭受质疑。隐私和安全等挑战可被视为技术挑战，持续聚焦这些能用技术手段控制并解决的技术挑战符合数字健康技术的开发人员和其他支持者的利益（Hedgecoe, 2010）。尽管我们承认这是对隐私和安全作为规范性问题的过度简化，但它与数字驱动的不平等和公共医疗保健系统中的企业俘获等问题形成的对比，说明了潜在规范性危害具有极大复杂性，而这些危害往往在伦理论辩中被掩盖或回避。本文的目标是，将数字健康的生命伦理学研究中最常见的问题，置于更广泛的社会、文化与政治现实背景之下加以讨论。这些社会、文化和政治背景，也正是这些数字健康问题产生的现实土壤。为实现这一目标，我们向科学技术学（STS）研究寻求理

论支持。

三、走向数字健康的社会技术伦理

（一）社会技术伦理与 STS

我们在论文中提出的关于数字健康伦理的社会技术进路直接来自 STS。STS 是一个跨学科研究领域，考察技术、科学与社会的相互关联与相互建构（Felt，Fouché & Miller et al. ，2017）。通过这种方式，STS 突出了人、实践、机构与其他物质现实，它们共同塑造了人们对科学技术及其影响的理解（Felt，Fouché & Miller et al. ，2017；Latour，2005；Mol，2003）。“社会技术”一词表明，诸如数字健康应用等与技术相关的问题不仅关乎物质技术本身，还关乎技术之间的相互依赖以及构建和使用它们的社会安排（Ananny，2016；Selbst，Boyd & Friedler et al. ，2019）。同样，从设计室的椅子和白板到调节人类互动的智能手机应用程序和视频会议软件，社会安排也总是被各种技术所渗透。因此，“社会技术”一词意味着将焦点从技术本身定义的问题扩展至更广泛的问题域，因为人们认识到技术是以一种既深刻塑造社会生活又为社会生活所塑造的方式建构并嵌入社会世界的（Verbeek，2006；Verbeek，2011）。

我们从安德鲁·塞布斯特（A. Selbst）等人的论文《社会技术系统中的公平性与抽象性》（“Fairness and Abstraction in Sociotechnical Systems”）中，提取出“社会技术系统”（sociotechnical system）这个概念。在分析中，塞布斯特等人概述了数据密集型技术应用中的一系列偏见是如何产生的。偏见并非技术本身所导致的，而是源于构建和内嵌于这些技术的社会和物质系统。在他们的工作之外，“社会技术系统”概念还有许多其他的理论来源，但出于本次分析的目的，我们指定其中的两个：基础设施研究和数字数据的政治经济学研究。

基础设施研究是 STS 领域的一种研究方法，它检视时常被忽略的物质基础，这些物质基础使日常生活成为可能（Star，1999；Edwards，Bowker & Jackson et al. ，2009）。聚焦基础设施的一个结果是，通过追溯特定分析点的延长线与延长的距离，揭示许多活动所依赖的基础设施之间的相互关联。以这种方式，通

过观察数字技术，我们得以了解它们在医疗健康领域所依赖的已联结的基础设施。

与之相关的是，数字数据的政治经济学研究揭示出数字健康技术收集、操纵和使用数据的典型的隐藏动机。在为该主题的一期特刊撰写的导言中，芭芭拉·普兰萨克（B. Prainsack）概述了数字数据的政治经济学研究是如何鼓励人们关注那些能够治理并使特定参与者从数据中开发价值的机构的（Prainsack，2020）。这一进路促使人们关注全球资本主义经济中的权力运作，这就要求地方机构与最新的资本主义趋势保持一致。通过这种方式，我们也关注塑造数字健康领域的更广泛的权力流动，并鼓励将其作为重要的规范性问题加以关注。

从技术到技术社会系统的视角拓展，引发了人们对那些以技术为中心的视角下可能被忽视的潜在伦理问题的关注。社会技术进路在技术的伦理分析中引入了一系列需要考虑的新的潜在伦理危害，从而对伦理分析产生更高层次的影响：它更明确地将我们的伦理注意力导向一个问题，即我们试图通过设计与部署特定技术实现一个怎样的世界。与简单地评估一开始就被确定为与伦理相关的一系列问题不同，社会技术进路允许人们探索与技术相距更远的问题，而这些问题可能也需要伦理关注；事实上，在这些更广泛的伦理问题得到承认之前，伦理分析——特别是就其对实现我们希望实现的那种世界的努力而言——被认为是不完整的。举例言之，研究者在设计数字健康技术时，最终可能还是绕不开分析特定司法辖区内与影响健康政策的盈利性技术公司相关的政策框架（Shaw，Rudzicz & Jamieson et al.，2019）。在这种情况下，健康政策问题可能对作为公共产品的健康系统结构产生重要影响，因而在与我们希望实现的世界相关的总的伦理分析中发挥重要作用。

在本节的其余部分，我们将概述可能产生数字健康伦理危害的社会技术领域的一般框架（表 1）。在下一节，我们将提出一种新的伦理分析进路，反思这些伦理危害并确定一条合乎伦理的前进道路。我们会提出从社会技术视角分析数字健康技术时可资参考的伦理危害分析框架。这一框架旨在为数字健康的社会技术伦理进路提供思路，研究者可借此了解我们可能在哪里找寻更广泛的伦理问题。框架无意全面涵盖与伦理相关的任一可行领域，而是试图在数字健康的社会技术系统的伦理分析中代表众多最突出的伦理考量。

表 1 数字健康伦理的社会技术进路的分析领域

分析领域	简要说明	伦理问题示例
应用软件	构成特定数字健康技术的代码行，以及它们促使或阻碍的与健康相关的实践	对边缘化群体执行精度较低的算法
材料设备与供应链	用于制造和分配设备（人类与数字健康技术通过这些设备互动）的实际材料	挖掘用于制造数字设备的原材料对环境造成负面影响
基础设施	数字健康运行所需的基础设施，包括诸如医疗保健提供者在提供虚拟护理时所处的建筑，使数字信号能够远距离传输的电缆和电线，以及提供数字通信的组织的公司结构等物质现实	生活在农村和偏远地区的人无法使用互联网
健康相关的个体实践	由于使用数字健康技术而被迫进行的活动和日常安排	持续的健康相关监测对心理健康产生不利影响
人际关系	数字健康技术能够通过塑造人们可获得的信息来源、期望和互动模式来影响人际关系	在社交媒体上分享与健康相关的错误信息对人际关系造成负面影响
组织政策	组织的结构、功能和日常程序，以及这些在组织规则和政策中的形式化方式	公司的员工监督制度
政府政策与监管俘获	由国家行为体制定的治理规则和进路，以及大型科技公司等数字健康利益相关主体对与健康相关的政府政策的影响	大型科技公司对健康相关政策的影响力持续增长

（二）一般框架

1. 应用软件

构成特定数字技术的代码行（lines of code），以及它们促使或阻碍的健康相关的实践，无疑具有重要的伦理意义。研究者围绕与健康相关的人工智能（AI）应用中的伦理问题已经做了大量工作，它们通常与决定某种数字健康技术功能的算法有关（Char，Abràmoff & Feudtner，2020；Morley，Machado & Burr et al.，2020）。尽管算法本身的透明度与公平性等问题很重要，但承认这些问题的伦理重要性与它们所属的更广泛的系统密切相关也是至关重要的（Selbst，Boyd & Friedler et al.，2019）。应用软件层面的伦理问题包括有效性、可用性、包容性、透明性，以及与数字健康产品的功能和直接使用相关的其他问题（Char，Abràmoff & Feudtner，2020；Shaw，Brewer & Veinot，2021）。

2. 材料设备与供应链

人类通过一定的设备(devices)与数字健康技术进行交互。用于构建和分配这些设备的实际材料(actual material),在伦理分析中往往被忽视,却与全面理解数字健康伦理高度相关。用于制造智能手机和其他数字设备的材料是从地球中开采出来并在国际运输的,其影响是加大对低收入国家低收入劳动力的剥削,使高收入国家的大公司受益(Hockenberry, 2021)。这种供应链所创建的系统以及高收入国家不断推进的数字消费循环,强化了地缘政治、结构性种族主义(structural racism)和气候危机(Hockenberry, 2021),从数字健康伦理的更广阔视野加强了它们的伦理相关性。尽管承认供应链和构成数字健康所需设备的材料的相关性给数字健康伦理带来了巨大挑战,但这并不意味着它们应该被排除在伦理分析之外。

3. 基础设施

正如前文概述的两个领域所论及的,数字健康有赖于硬件和软件。然而,数字健康也依赖于一系列不同类型的基础设施(infrastructure)。这些基础设施,包括提供虚拟护理的工作人员的办公场所、使数字信号得以远距离传输的电缆和电线,以及提供数字通信的组织的企业结构(Østerlund & Bjørn, 2011; Gray, Gerlitz & Bounegru, 2018)。基础设施可能对数字健康产生至关重要的伦理影响,如缺乏高速互联网可能会将某个特定社群排除在数字医疗保健之外(Shaw, Brewer & Veinot, 2021)。

4. 健康相关的个体实践

数字技术被用于各种与健康相关的应用,其中许多应用旨在促进个体的健康活动和疾病管理(Bayram, Springer & Garvey et al., 2020)。数字健康技术通常融入了自我追踪机制,其影响在于能鼓励人们监管自身的行为和习惯,这意味着他们对自己的行动是否以及如何与预期的社会规范保持一致有了更高的认识(Lupton, 2017b; Felt, 2015)。尽管这种自我监管的后果可能包括改善健康和预防疾病,但就自我追踪和"助推"(nudge)技术在塑造和限制人类行为上的影响而言,仍有更广泛的问题等待提出(Yeung, 2017)。技术对个体心理健康(mental well-being)的影响与对个体行为(individual actions)的影响均与伦理相关,应该在数字健康的相关伦理分析中得到承认。

5. 人际关系

数字健康技术能以各种方式影响人际关系。其中一个例子是,社交媒体应用对公众理解与健康相关的科学与政策存在显著影响(Allington, Duffy & Wessely et al., 2020; Schillinger, Chittamuru & Ramírez, 2020)。社交媒体在与健康相关的应用上,有可能建立人际网络,强化网络用户对健康相关问题的特定认知,对公共健康造成潜在的破坏性影响。另一个例子是经技术中介的沟通对医患关系的影响(Oudshoorn, 2011)。尽管数字健康与医患关系的确切关联尚不清楚,但面对面的护理与数字介导的护理之比较仍然与数字健康在伦理上相关。

6. 组织政策

数字技术具有极大地重塑日常工作实践,从而重塑组织结构与功能的潜能(Barley, 1986)。与健康相关的组织从实体环境向数字工作环境转变的方式,可能对医疗卫生工作的性质和患者照护(patient care)的性质产生重要影响(Shaw, Agarwal & Desveaux et al., 2018; Shaw, Rudzicz & Jamieson et al., 2019)。医疗保健系统的运作方式与公共利益十分相关,扩大了被认为与数字健康的伦理分析相关的伦理问题范围。一个值得一提的要点是诸如保险公司此类组织的影响,这些组织使用数字健康技术来收集有关个体行为的信息并据此塑造其产品(McFall, Meyers & Hoyweghen, 2020)。这些实践由于数字健康技术的进步而获得新的效能,它们在未来保险行业应该发挥的作用是一个需要密切伦理关注的组织政策问题。

7. 政府政策与监管俘获

在企业不断加码投资收集与分析大量健康相关数据的背景下,政府监管变得尤为重要。欧盟的《通用数据保护条例》(GDPR)作为处理健康相关数据的数据保护法的最新进展之一,代表朝向更全面的公共保护(public protection)迈出的重要一步。然而,数字健康技术的快速发展及其制造商的组织实践对 GDPR 的有效性提出了重要挑战。譬如,卢卡·马瑞利(L. Marelli)等人指出了 GDPR 未能有效覆盖的一系列数字健康实践,包括新企业主体日益扩大的影响力、加强医疗保健和生活方式之间的联系、增加对预测分析的依赖,以及社会分类将技术用户划分为不同群体等(Marelli, Lievevrouw & Van Hoyweghen, 2020)。超出现有政策对数字健康实践覆盖范围的,是这些企业主体对医疗保健系统的战略

及运行的持续增长的影响力。盈利性科技公司越来越多地进入数字健康领域，突出了对科技公司和医疗保健系统彼此冲突的动机投以伦理关注的迫切需要（Sharon，2020a；Sharon，2020b）。

沿着数字健康伦理的社会技术进路，上述分析领域中存在的显著伦理危害可能被识别出来，这些领域代表对那些在传统上与生命伦理学相关的有限视角的背离（Hedgecoe，2004；Hedgecoe，2010）。社会技术进路鼓励伦理研究者参与必要的工作，以便更清楚地理解数字健康技术在不同领域的具体应用带来的伦理问题。此种努力可能需要回顾有关这些主题的社会科学文献或进行新的实证研究，以揭示特定技术的含义及其生产与分配方式。然而，在确定和理解这种技术的潜在危害之后，如何在这些危害和预期的益处之间做出权衡并得出有意义的伦理结论？接下来我们将讨论这个重要的问题。

四、迈向我们期待的世界

（一）伦理视角拓展

如果说数字健康伦理的社会技术进路带来的第一个重要转变在于拓宽了理应考虑的问题的范围，那么它带来的第二个重要转变则是聚焦于世界建构。与现象学中的“世界建构”（world-building）概念相反，我们所说的“世界建构”是指通过实践和制度对创造特定类型的世界所做出的分布式贡献，这些实践和制度使特定技术的可持续发展成为可能。这种对世界建构的理解，一方面在更广泛的层面与 STS 领域的文献相一致，这些文献关注构成创新的多元活动场景；另一方面与批判政治经济学（critiacl politiacl economy）进路的研究方法相契合，这一方法探讨特定的创新是否有助于建构我们希望实现的世界（Prainsack，2020；Costanza-Chock，2020）。社会技术进路令更多人意识到，技术的创造与传播能够以与伦理相关的方式影响世界，并且这种影响是多方面、多途径的。我们认为，关注范围如此广泛的一系列议题，要求研究者对技术的后果及其建构方式所带来的世界具有总括性的了解。从社会技术伦理的视角对某项技术做出综合评估，离不开对该技术助力创造的世界，以及谁在此世界中受益或处于不利地位的

理解。这一进路优先考虑社会正义。

将个体的伦理视角扩展至社会技术系统中的诸多要素,有助于人们更全面地理解社会技术系统的规范意义。在这一更广泛的伦理分析层面,我们建议将伦理关注最自然地集中于两个方面:一是世界建构,二是支持人人享有社会正义的价值承诺。在追溯社会技术系统中的联系时,原本被认为没有联系的共同体之间的相互联系就会浮现在眼前,它们之间的相互依存关系在伦理上变得更加突出了。与公共健康伦理的最新进路一致,社会技术伦理要求对地方的、国家的、全球的利益与责任之平衡予以伦理关注(Benatar, Upshur & Gill, 2018)。这种对社会正义的关注,源自同受特定数字健康技术影响的互联共同体对彼此的团结承诺;也正是这种对社会正义的关注,体现出数字健康的社会技术伦理的鲜明特点。遵循这一进路,需要借鉴生命伦理学相关领域中伦理治理的常见方法。我们将借用“预期治理”(anticipatory governance)的概念来描述其中的两种方法。

(二)社会技术伦理的公众参与和预见

数字健康伦理的社会技术进路与“预期治理”的概念存在强烈共鸣(Guston, 2014)。大卫·古斯顿(D. Guston)将“预期治理”定义为:“一种在社会中广泛延伸的能力,它可以利用各种投入来管理新兴的知识型技术,并且这种管理是可能实现的。”(Guston, 2014)这一定义将“预期治理”概念化为一种分布式实践,在特定社会的创新功能中被制度化。尽管与预期治理相比,数字健康伦理的社会技术进路具有更加温和的目标,即提供更直接的伦理分析,但它确实从预期治理中得出两个相互关联且各自重要的见解——“公众参与”(engagement)和“预见”(foresight)。

从预期治理得出的第一点启示是,非专业公众的参与很重要,他们能为技术的意义和可期待性评判提供不同视点(Guston, 2014; Barben, Fisher & Selin et al., 2008)。这种观点与假设的交流,能帮助伦理专家更好地理解道德预设(moral assumptions)以及不同公众持有的与特定技术相关的知识和信念。公众参与旨在确保在创新、政策与技术方面代表性不足的观点能被纳入讨论的范畴,并对伦理问题的建构产生影响。然而,公众参与同样面临挑战——如何在公众

中确定参与者并以有意义的方式确保资源安全，需要承诺行动（committed action）和能够提供支持的、资源充足的环境。

来自预期治理的第二点启示是预见的重要性，这在新兴技术的相关研究中已被详细论及（Floridi & Strait, 2020；Moor, 2005；Lucivero, 2016）。在关于预期治理的正式文献中，预见在其完全意义上被纳入一种识别当前趋势并想象各种潜在未来的可能性与重要性的多元方法实践（Barben, Fisher & Selin et al., 2008）。这些在与相关公众协商中被确定的潜在未来能够为当前的治理决策提供方法。这对数字健康伦理的社会技术进路而言，则是预见从各种社会技术分析领域辨认出的更广泛的伦理问题的潜在影响。其目的是为决策提供更清晰的理解基础，包括清楚地了解焦点数字健康技术所倡导的那种世界，以及清楚地理解受到这些技术影响的互联共同体将要面对的后果。

在此所阐述的方法既依赖多元视角的参与，也离不开对更理想的未来的表达。两者都不可能是完美的，因而伦理分析总是片面和不完整的。这不是伦理分析的“次要”（lesser）版本，而是社会技术视角下唯一被视为可能的伦理形式。其目的在于，分析不同贡献者对某一特定技术的影响的规范性观点，进而评估其对未来的影响。该方法旨在通过聚焦整个社会技术系统中涉及数字健康技术开发与分配的共同体的社会正义，为所有人建立一个更美好的世界。

我们在此阐述的方法对健康系统决策层的实际意义与上述两点启示相关。当健康系统打算采用新技术时，它们可以加入公众参与的系统过程，以便建立一个富有意义且为不同公众所接受的流程与治理方式，它将包括但不限于那些在结构上被边缘化的人。这种方法使健康系统能够发现那些带薪从事采购与实施数字健康技术的工作人员可能无法理解的问题。

此外，健康系统的领导者可以采取一种方法，明确预见采用技术对健康系统利益相关者的潜在负面后果。基于公众参与带来的各异见解可知，这种潜在的负面后果不仅会影响使用该技术的临床医生，还会改变行政人员、领导者、患者和其他参与构建、分配和管理技术的利益相关者在系统中的角色。

最后，卫生系统的领导者可以向熟悉数字健康技术特有伦理问题的伦理专家寻求具体的意见。尽管这是一个新兴领域，但随着数字健康技术在全球健康系统中发挥日渐突出的作用，关于数字健康技术伦理的具体咨询将变得越来越

重要。我们于文中阐述的更广泛的视角,以及在此说明的实际意义,均有助于促进未来数字健康技术可持续的、合乎伦理的应用。

五、结论

我们在文中所提出的数字健康的社会技术伦理,建基于传统伦理研究进路的认识论批判与规范性基础批判。根据这种批判,我们提出一种观点,提请读者关注由特定数字健康技术所涉及的社会技术系统和与之相关的众多共同体的福祉这两方面所代表的更广泛的问题。当规范的关注转向这些共同体的福祉时,团结的价值以及对社会正义的承诺在伦理分析中变得更加突出。重点转变成建立一个对所有人都更好的世界,而非一个只对少数特权利益相关者更好的世界。

我们于文中表明,不仅技术本身需要伦理关注,令技术得以应用并由技术所创造的世界同样需要伦理关注。正如塞布斯特等人所说:“公平和正义是就业和刑事司法等社会和法律系统的属性,而非其中技术工具的属性。”(Selbst, Boyd & Friedler et al., 2019)数字健康伦理的研究者,需要根据这一基本观点为医疗保健和公共健康领域新兴技术的伦理分析制定最合适的策略。

参考文献

Allington, D., B. Duffy & S. Wessely et al. 2020, “Healthprotective Behaviour, Social Media Usage and Conspiracy Belief During the COVID-19 Public Health Emergency.” *Psychological Medicine* 51(10).

Ananny, M. 2016, “Toward An Ethics of Algorithms: Convening, Observation, Probability, and Timeliness.” *Science, Technology, & Human Values* 41(1).

Auerbach, A. 2019, “Evaluating Digital Health Tools: prospective, Experimental, and Real World.” *JAMA Internal Medicine* 179(6).

Barben, D., E. Fisher & C. Selin et al. 2008, “Anticipatory Governance of Nanotechnology: Foresight, Engagement, and Integration.” in E. Hackett, O. Amsterdamska & M. Lynch et al. eds., *The Handbook of Science and Technology Studies*. Cambridge, MA.: The MIT Press.

Barley, S. 1986, "Technology as An Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments." *Administrative Science Quarterly* 31(1).

Bayram, M., S. Springer & C. Garvey et al. 2020, "COVID-19 Digital Health Innovation Policy: A Portal to Alternative Futures in the Making." *Omics* 24(8).

Benatar, S., R. Upshur & S. Gill 2018, "Understanding the Relationship Between Ethics, Neoliberalism and Power as A Step Towards Improving the Health of People and Our Planet." *The Anthropocene Review* 5(2).

Brall, C., P. Schröder-Bäck & E. Maeckelberghe 2019, "Ethical Aspects of Digital Health from A Justice Point of View." *European Journal of Public Health* 29.

Char, D., M. Abràmoff & C. Feudtner 2020, "Identifying Ethical Considerations for Machine Learning Healthcare Applications." *American Journal of Bioethics* 20(11).

Costanza-Chock, S. 2020, *Design Justice: Community-Led Practices to Build the Worlds We Need*, Cambridge, MA.: The MIT Press.

Edwards, P., G. Bowker & S. Jackson et al. 2009, "Introduction: An Agenda for Infrastructure Studies." *Journal of the Association for Information Systems* 10(5).

Felt, U. 2015, "Sociotechnical Imaginaries of 'the Internet', Digital Health Information and the Making of Citizen-Patients." in S. Hilgartner, C. Miller & R. Hagendijk eds., *Science and Democracy: Making Knowledge and Making Power in the Biosciences and Beyond*, New York: Routledge.

Felt, U., R. Fouché & C. Miller et al. 2017, *The Handbook of Science and Technology Studies*, Cambridge, MA.: The MIT Press.

Floridi, L., A. Strait 2020, "Ethical Foresight Analysis: What It Is and Why It Is Needed?" *Minds and Machines* 30.

Fox, R. & J. Swazey 2005, "Examining American Bioethics: Its Problems and Prospects." *Cambridge Quarterly of Healthcare Ethics* 14(4).

Fox, R. & J. Swazey 2008, *Observing Bioethics*, London: Oxford University Press.

Fox, R. & J. Swazey 2010, "Guest Editorial: Ignoring the Social and Cultural Context of Bioethics Is Unacceptable." *Cambridge Quarterly of Healthcare Ethics* 19(3).

Fox, R. 2016, "Moving Bioethics Toward Its Better Self: A Sociologist's Perspective." *Perspectives in Biology and Medicine* 59(1).

Gray, J., C. Gerlitz & L. Bounegru 2018, "Data Infrastructure Literacy." *Big Data & Society* 5 (2).

Greenhalgh, T., R. Procter & J. Wherton et al. 2012, "The Organising Vision for Telehealth and Telecare: Discourse Analysis." *BMJ Open* 2(4).

Guston, D. 2014, "Understanding 'Anticipatory Governance'." *Social Studies of Science* 44.

Haluza, D., M. Naszay & A. Stockinger et al. 2016, "Prevailing Opinions on Connected Health in Austria: Results from An Online Survey." *International Journal of Environmental Research and Public Health* 13(8).

Hedgecoe, A. 2004, "Critical Bioethics: Beyond the Social Science Critique of Applied Ethics." *Bioethics* 18(2).

Hedgecoe, A. 2010, "Bioethics and the Reinforcement of Socio-Technical Expectations." *Social Studies of Science* 40(2).

Henwood, F. & B. Marent 2019, "Understanding Digital Health: Productive Tensions at the Intersection of Sociology of Health and Science and Technology Studies." *Sociol Health Illness* 41(S1).

Hockenberry, M. 2021, "Redirected Entanglements in the Digital Supply Chain." *Cultural Studies* 35(4-5).

Joerin, A., M. Rauws & R. Fulmer et al. 2020, "Ethical Artificial Intelligence for Digital Health Organizations." *Cureus* 12(3).

Latour, B. 2005, *Reassembling the Social: An Introduction to Actor-Network-Theory*, London: Oxford University Press.

Latulippe, K., C. Hamel & D. Giroux 2017, "Social Health Inequalities and E-Health: A Literature Review with Qualitative Synthesis of Theoretical and Empirical Studies." *Journal of Medical Internet Research* 19(4).

Lucivero, F. 2016, *Ethical Assessments of Emerging Technologies*, Cham: Springer.

Lupton, D. 2017a, *Digital Health: Critical and Cross-Disciplinary Perspectives*, New York: Routledge.

Lupton, D. 2017b, *Wearable Devices: Sociotechnical Imaginaries and Agential Capacities*, Cambridge, MA.: The MIT Press.

Marelli, L., E. Lievevrouw & I. Van Hoyweghen 2020, "Fit for Purpose? The GDPR and the Governance of European Digital Health." *Policy Studies* 41(1).

Marent, B. &F. Henwood 2021, "Digital Health." in K. Chamberlain &A. Lyons eds., *Routledge International Handbook of Critical Issues in Health and Illness*, New York: Routledge.

Mattingly, C. &J. Throop 2018, "The Anthropology of Ethics and Morality." *Annual Review Anthropology* 47(1).

Mattingly, C. 2012, "Two Virtue Ethics and the Anthropology of Morality." *Anthropological Theory* 12(2).

McFall, L., G. Meyers &I. Hoyweghen 2020, "Editorial: The Personalisation of Insurance: Data, Behaviour and Innovation." *Big Data &Society* 7(2).

Milosevic, Z. 2019, "Ethics in Digital Health: A Deontic Accountability Framework." 2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC).

Mol, A. 2003, *The Body Multiple*, Raleigh: Duke University Press.

Montagni, I., T. Cariou &T. Feuillet et al. 2018, "Exploring Digital Health Use and Opinions of University Students: Field Survey Study." *JMIR mHealth and uHealth* 6(3).

Moor, J. 2005, "Why We Need Better Ethics for Emerging Technologies." *Ethics and Information Technology* 7.

Morley, J., C. Machado &C. Burr et al. 2020, "The Ethics of AI in Health Care: A Mapping Review." *Social Science &Medicine* 260(C).

Mulvenna, M., R. Bond &J. Delaney et al. 2021, "Ethical Issues in Democratizing Digital Phenotypes and Machine Learning in the Next Generation of Digital Health Technologies." *Philosophy &Technology* 34(4).

Murray, E., E. Hekler &G. Andersson et al. 2016, *Evaluating Digital Health Interventions: Key Questions and Approaches*, Washington: Elsevier.

Myser, C. 2003, "Differences from Somewhere: The Normativity of Whiteness in Bioethics in the United States." *American Journal of Bioethics* 3(2).

Nebeker, C., J. Torous &R. Ellis 2019, "Building the Case for Actionable Ethics in Digital Health Research Supported by Artificial Intelligence." *BMC Medicine* 17(1).

Østerlund, C. &P. Bjørn 2011, "Socio-Material Infrastructure in Emergency Departmental Work." Infrastructures for Healthcare: Global Healthcare. Proceedings of the 3rd International Workshop on Infrastructures for Healthcare.

Oudshoorn, N. 2011, *Telecare Technologies and the Transformation of Healthcare*, doi: 10.

1057/9780230348967.

Prainsack, B. 2020, "The Political Economy of Digital Data: Introduction to the Special Issue." *Policy Studies* 41(5).

Ruotsalainen, P. & B. Blobel 2020, "Health Information Systems in the Digital Health Ecosystem: Problems and Solutions for Ethics, Trust and Privacy." *International Journal of Environmental Research and Public Health* 17(9).

Schillinger, D., D. Chittamuru & A. Ramírez 2020, "From 'Infodemics' to Health Promotion: A Novel Framework for the Role of Social Media in Public Health." *American Journal of Public Health* 110(9).

Schmietow, B. & G. Marckmann 2019, "Mobile Health Ethics and the Expanding Role of Autonomy." *Medicine, Health Care and Philosophy* 22(4).

Selbst, A., D. Boyd & S. Friedler et al. 2019, "Fairness and Abstraction in Sociotechnical Systems." doi: 10.1145/3287560.3287598.

Sharon, T. 2020a, "Blind-Sided by Privacy? Digital Contact Tracing, the Apple/Google API and Big Tech's Newfound Role as Global Health Policy Makers." *Ethics and Information Technology* 23.

Sharon, T. 2020b, "Beyond Hostile Worlds: The Multiple Sphere Ontology of the Digitalization and Googlization of Health." doi: 10.2139/ssrn.3633371.

Shaw, J., F. Rudzicz & T. Jamieson et al. 2019, "Artificial Intelligence and the Implementation Challenge." *Journal of Medical Internet Research* 21(7).

Shaw, J., L. Brewer & T. Veinot 2021, "Health Equity and Virtual Care: A Narrative Review of Recommendations Arising from the COVID-19 Pandemic." *JMIR Formative Research* 5(4).

Shaw, J., P. Agarwal & L. Desveaux et al. 2018, "Beyond 'Implementation': Digital Health Innovation and Service Design." *NPJ Digital Medicine* 1.

Shaw, J., T. Jamieson & P. Agarwal et al. 2018, "Virtual Care Policy Recommendations for Patient-Centred Primary Care: Findings of A Consensus Policy Dialogue Using A Nominal Group Technique." *Journal of Telemedicine and Telecare* 24(9).

Star, S. 1999, " The Ethnography of Infrastructure." *American Behavioral Scientist* 43(3).

Vayena, E., T. Haeusermann & A. Adjekum et al. 2018, "Digital Health: Meeting the Ethical and Policy Challenges." *Swiss Medical Weekly* 148.

Veinot, T., H. Mitchell & J. Ancker 2018, "Good Intentions Are Not Enough: How Informatics Interventions Can Worsen Inequality." *Journal of the American Medical Informatics Association* 25(8).

Verbeek, P.-P. 2006, "Materializing Morality: Design Ethics and Technological Mediation." *Science, Technology, & Human Values* 31(3).

Verbeek, P.-P. 2011, *Moralizing Technology: Understanding and Designing the Morality of Things*, Chicago: University of Chicago Press.

WHO 2016, *Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment*, Geneva: World Health Organization.

Winters, N., S. Venkatapuram & A. Geniets et al. 2020, "Prioritarian Principles for Digital Health in Low Resource Settings." *Journal of Medical Ethics* 46(4).

Yeung, K. 2017, "'Hypernudge': Big Data as A Mode of Regulation by Design." *Information, Communication & Society* 20(1).

数字资本主义和智能技术的当代批判

——评《过度智能》

周银知*

摘要:人类正处于探索数字化道路的险途,数字化和智能化成为数字时代的典型特征,并以前所未有的深度和广度捕获了个体、家庭和城市,数据成为一种商业模式。美国青年学者贾森·萨多夫斯基新著《过度智能》从政治学和哲学角度,对当前数字资本主义和智能技术的发展进行了犀利批判,指出了技术由谁主导、为谁服务的权力和利益问题,凝结了作者对理解、对抗和挑战西方技术政治体系的深刻观察。

关键词:数字资本主义　智能技术　技术政治

当今,对数字资本主义与智能技术议题的讨论热火朝天。人类正处于探索数字化道路的险途,数字化和智能化成为数字时代的典型特征,并以前所未有的深度和广度捕获了个体、家庭和城市,数据成为一种商业模式。美国青年学者贾森·萨多夫斯基(J. Sadowski)新著《过度智能》于2020年付梓,中国传媒大学新媒体研究院副研究员徐琦于2022年将其翻译引介至国内。正如作者在"余论"中所言,他在书中展示了智能社会的隐秘危害以及危害是如何扩散和日益严重的,此书正是他不遗余力地对现有秩序所做的批判(萨多夫斯基,2022:233)。

《过度智能》从政治学和哲学角度对当前数字资本主义和智能技术的发展进行了犀利批判,指出了技术由谁主导、为谁服务的权力和利益问题,凝结了作者对理解、对抗和挑战西方技术政治体系的深刻观察,是我们了解数字资本主义政治体系运作机制的一本很好的参考书。基于此,本文分为四部分:第一部分梳理《过度智能》每个篇章的主要内容和观点;第二部分以书中的部分核心观点为

* 周银知,武汉理工大学马克思主义学院。

支撑,深入分析数字资本主义的新型剥削和控制形态并对此展开批判;第三部分梳理学术界对技术的批判理路,并与书中的观点展开对话,进一步挖掘书中的深层次内涵和逻辑;第四部分结合中国式现代化进程,对作者解构资本的新范式所提出的路径进行评析和借鉴。

一、数字资本主义的要务:收集与控制

《过度智能》一书由三部分组成:第一部分(第1—3章)主要为数字资本主义的理论问题,探讨数据的收集、控制和资本化;第二部分(第4—6章)分别探讨了规模递进的三种智能场域,即智能自我、智能家居、智能城市的发展和问题;第三部分(第7章)是在上述两个部分对智能技术的资本进行解构之后,提出建构方式和可能的现实路径。

第1章"数据世界"指明,数据作为资本形式被收集、应用和生产出来,成为数字化时代资本主义发展的核心要素。数据已然成为公司经营的重要商业模式,甚至各个行业都将数据化作为主要的业务予以推进。数据价值体现为可用于分析和追踪定位、优化系统、管理事务、建立概率模型、制造产品、实现资产增值等。智能技术一方面成为制造数据世界的手段,另一方面也依赖源源不断的数据以保持更新和运转。数字资本主义通过对所有人和所有物尽可能多地持续性收集数据,创建起了监视、管理和控制人的社会系统。

第2章阐释了人类社会发展的三种权力形态:武力、规训和控制。作者认为,武力在封建社会体现为王权和君权,而在人类的历史长河中,更为普遍地体现为警察、父母等角色掌握的威胁和惩罚的权力(萨多夫斯基,2022:48)。在此,武力可以作为权力的第一等级。"规训"是福柯(M. Foucault)提出的概念,与资产阶级的发展密切相关。它倡导特定的生活方式、思维方式和行为方式,使得人们将这些信条、习惯和规范内化。在这种形态下,权力便决定了如何规训、建立何种规训制度和方式,这是对人的生命权利的技术运用。因此,规训更为高级,可以作为权力的第二等级。控制被作为权力的第三等级,作者认为,控制不是用暴力威胁人们,也不是用规训的方式来塑造人们,而是通过设定参数和建立检查站来规范人们的行为和发挥作用,人们往往感受不到所处的控制系统(萨

多夫斯基,2022:49),其象征便是计算机网络。德勒兹(G. Deleuze)的控制理论对理解智能技术系统的运转及其对人类生活的影响具有重要的启发作用。在万物互联的社会中,智能技术的根茎已经遍布整个社会,并不断复制和蔓延。数据世界中的每一个体都是自我的“分体”,即独立的个体被分解为身体、行为、身份、特征、事件等部分而分别存在。人们如果想进入各类软件系统,就需要通过用户许可协议,提供自身的数据,比如密码、面部特征、指纹等。在现实生活中,这三种权力形态交织、融合和共生,共同形成了人类发展过程中的支配系统。第 1—2 章说明,智能技术的发展是由数字资本主义的收集、控制两大要务所驱动的。换言之,数字资本主义的发展依托于智能技术发展,对所有人和所有物进行数据收集,并对所有人和所有物实施控制。

第 3 章论述了数字资本主义运转中的十大论题。作者认为,数据化和数据收集意味着盗窃和暴力,其关键在于如何使用抽象化了的人的身体、行为和事件的数据,使用过程中的目标设定、价值选择和利益考量体现了数字资本的偏见与剥削性(萨多夫斯基,2022:76)。例如,大型互联网平台实际上是数字资本主义的“新房东”,用户则通过默认平台许可协议和放弃个人数据所有权来支付“租金”,数据所有权和使用权被转移至平台;技术官僚意识形态的核心是一种“解决主义”,将世界上的问题看作有待技术解决的问题,只要技术提供了合理化的解决方案,世界便可以按照技术设计的图景来创造。

第 4—6 章中,作者通过现实生活的丰富案例描述了智能技术对个体自我、家庭生活空间和城市空间的把握与控制(萨多夫斯基,2022:15)。涉及的问题主要分为三个方面:一是控制权力问题,即谁拥有数据和实施控制行为,这涉及为谁服务的利益之争,技术的掌舵者或将成为权力的拥有者;二是处在超级全景监狱下的人的生存境遇,被数据捕获的个体自我、家庭生活空间和城市空间最终成为技术的控制物;三是数字编码体系中的公平、民主和隐私问题等凸显。这三章集合了作者对当前智能社会三大主体场域的现实揭露和尖锐批判。

第 7 章集中论述了作者提出的新建构方式,包括“离经叛道”地废止那些没有实际改善生活和造福社会的技术、使创新过程民主化、进行数据监管和数据集体化。作者使用牛顿第三定律的隐喻来表明其反对技术垄断的信心(萨多夫斯基,2022:201)。牛顿第三定律指出:两个物体之间的作用力和反作用力总是大

小相等,方向相反,作用在同一条直线上。作者借此表示反对技术政治统治的集体行动作为反作用力正在形成团结和批判的力量,其力量的强度和影响力有望与技术霸权和技术垄断的作用力相匹敌。

在资本适应数字化、智能化的时代,作者对数字资本主义和智能技术所要实现的收集与控制两大要务的观察深入现实、切中肯綮。他认为,数字资本主义的创新精神让智能技术日臻完善,实现了对个体、家庭和城市等所有人和所有物的抽象化、数据化和再编码。我们确实在生活和工作中已经感受到了这样的现实,即人的方方面面无不被数据化,也无时无刻不被数据所引导和控制。同时,该书鲜明地提出了技术政治问题,将问题的实质直指谁控制技术和数据、为谁服务、为何如此的根本性问题。智能技术问题首先应该是政治经济问题,应当关注技术的政治和经济环境,作者由此揭示了美国的技术政治现实。因此,该书的意义不仅在于传递了对数字资本主义发展的学术之思,还在于对"批判的武器"的应用和发挥,书中大量的案例分析也反映了作者对学术源于现实并观照现实的期盼。智能技术已经成为全球发展趋势,与我们大家息息相关。在这样的现实和语境下,探讨智能技术发展的工具理性和价值理性,不仅是直面数字资本主义发展的压迫和剥削问题,更是关于人与技术之间以及未来社会的理想图景的根本性问题。作者尽力对数字资本主义政治经济体系进行批判,揭露智能社会隐秘的技术政治体系真相,并积极建构干预、对抗和挑战技术政治体系的路径。

二、数字资本主义:"新房东"和"分体"

"数字资本主义"是一个宽泛的用语,作者在这里主要指向数字技术适应资本化运作从而产生对智能技术的创新驱动,以及数据作为资本要素被资本化的现象。作者鲜明指出,平台就是数字资本主义的"新房东",并在书中给出了"智能"的定义,即"智能"意味着一个物体被嵌入了用于数据收集、网络连接和增强控制的数字技术,这显示了数字资本主义和智能技术中新的剥削形式和控制路径(萨多夫斯基,2022:2)。数字资本主义的"新房东"指当代众多行业的大型互联网平台,它们真正收取的"租金"不仅包含直观上的"软件即服务"的业务收入,还体现为掌握了由海量用户产生的数据的所有权。其中,前者以后者源源不

断的数据输入为基础来保持运转,后者实际上成为平台的源头活水。因此,影响“新房东”“租金”收入的关键因素是数据的来源问题,只要平台掌握了持续的活跃数据,软件服务和数据再编码便拥有了稳定的动力。

具体来说,“新房东”的概念体现了作者对数字资本主义批判的两大要点。一是以饱含正义的谴责之态深刻揭示了数字平台的剥削性,作者用“阴险”“暴力”“排挤”“榨取”“盗窃”这类富有情感的词描述资本化的数字运作,将智能技术的应用作为环境的产物,强调层层代码之下隐藏的是一系列的偏见、规范和假设。数字平台收集数据忽视了或简化了用户知情权和补偿权的问题,这是对用户个人信息和用户多元数据的偷窃。“新房东”与传统房东都是食利者,不同的是随着商业模式的更新,“新房东”不再对所有权收费,而是通过搭建“共享”“服务”的数字平台建立自由和平等的幻象,通过用户许可协议、软件许可协议等方式享有软件及其产生的数据的所有权。由于用户许可协议通常是规范性和强制性的,用户只有默认或同意协议才能够使用相关软件和硬件,因此平台便自然而然地获得了对软件进行访问、控制和提取数据的合理权利。实际上,平台冠以“分享”“互联”之名,却将用户置于不公平、不合理和丧失所有权的境地。正如其他学者所言:“数字资本主义凭借数据和平台,不仅维持和延续了已有的竞争优势和垄断地位,而且通过营造一种自由参与、民主双赢和信息共享的氛围实现更隐蔽的资本增殖和更深层的劳动盘剥。”(叶泽雄、鄢然,2023)这些观点在学界已成为共识。往前一步,数字平台用户的劳动在“新房东”语境下获得了类似农民在土地上耕种般的根基,即用户在数字平台这块土地上进行的信息、注意力、数据等耕作理当被视为数字劳动的一种形态,应当给予劳动肯定。这为我们辨析数字劳动的资本逻辑和用户劳动的非正义性提供了有益观点。

二是直言技术的政治性。“新房东”简洁透彻地指明了数字平台资本相较于封建社会在当今数字社会性质不变的剥削特征,剥削的背后是政治的角逐。作者认为,互联网平台形成的数据信息和数据系统以第三者的视角突出了其公平、客观和中立的特性,掩盖了数据收集、设计和运用中的目标设定、价值观、利益、制度等属人的主观性因素(萨多夫斯基,2022:105)。在此,值得追问和质疑的问题是:数据是客观和中立的吗?数据是衡量我们行为的标准吗?作者警示到,智能社会的建构是一项政治项目,是关于人们想象力的战争(萨多夫斯基,

2022:87)。蓝江认为:“我们通过数字之门,通过一个注册的身份,成为资本主义数据猎手的猎物,我们在虚无的浏览和刷屏中,不仅让自己脱离于大地,也让自己成为数字社会资本主义机器的一部分。”(蓝江,2023)事实上,大型互联网平台凭借强大的规模优势和行业主导地位,在对人们的行为引导、修正、规范和控制上起着无法抗拒的作用。个体往往将平台作为独立第三方来看待,忽视了平台所呈现的结果并非对世界的正确反映和判断。我们需要对数据和信息的完整性与真实性持批判的态度,至少不是完全被数据系统决定什么是正确的、好的、有价值的和值得追求的。这对反思数字平台带有的意识形态欺骗性、智能技术的社会建构和政治性具有强烈的启示意义,而且在资本主义政治经济体系中,怎么强调技术的政治性都不为过,这反而对数字技术的设计、开发和应用等环节的政治性实质进行了理性的揭露。

除此之外,作者论述的“分体”思想是上述“新房东”概念的重要理论支撑。构成“新房东”的“租客”是每个个体或被部分化的“分体”,组成“新房东”进行生产的要素是个体或“分体”提供的源源不断的具有实质内容的数据和信息,使得“新房东”贯彻实施精细化微观管理成为可能的便是数据最大化原则之下“分体”式的存在。个体不断被“分体”化、原子化的过程,实际上是平台实施数据最大化原则的有效方式和必然结果。数据采集将人们的不同属性和行为作为收集目标,智能技术便会实现从周围环境中剥离出来提取特定要素来完成数据收集的任务。独立的个体不断被身体、行为、事件等部分数据“分体”化,作为不连续的、分散的、部分的“分体”存在于数字世界中。同时,智能技术也增强了通过属性、行为或类别来代表一个个体的能力。个体是复杂、矛盾和动态的存在,抽象、分解、提取便成为理解人、改变人和控制人的高效方式。这导致的严重后果便是数字围场中微观管理至死,即基于对人们各方面数据最大化的收集进而实现微观层面的极致控制。数据的爆炸式增长和使用,强化了第二次工业革命以来的绩效管理思想。随着对个体、城市和社会层面数据收集、分析、储存和运用的能力不断加强,公司、政府和社会获取的信息越来越多,便越有可能去建立有效的量化标准和指标,不断改进和完善工作流程、供应链管理甚至城市建设。在此,可以将个体、城市、社会在数据中的“分体”式存在看作在智能社会中进行微观细化管理的理论深思和现实依据。

在实践层面,“分体”式存在也为微观管理在智能社会中的实施提供了可行的思路和方案。数据化的个体以“分体”式存在,并被分析活动进行审视、检查、评估。作者指出,数据代理商最大化收集所有人的数据和信息,将其细分为不同的市场,进行标签化、分类化处理,并提供信用评级服务(萨多夫斯基,2022:103)。例如,保险公司会根据人们的信用卡消费记录、健身会员卡资格和健身次数等决定保费。微观管理不仅体现在对个体的评分、分类上,也体现在城市和社会层面的微观控制上。尤其是,遥感技术、互联网、计算机技术、算法和算力的发展让整个城市都处于监控器的视觉范围内。所有人、所有物均以数据化的形式部分地存于数据系统之中,管理者也借助数据系统建构、影响和改变代表个体的“分体”式存在。

那么,值得深思的是:被智能技术数据化和“分体”化的人能否代表真正的人?这种做法是以什么样的价值观来审视人的,是以什么样的世界观来构序世界的?作者指出,抽象化的数据能够形成对世界排序和建构的强大力量,这种探知世界特征和动态的方式能够使世界分类化、规范化、清晰化,但排除了审视世界的其他方式(萨多夫斯基,2022:75)。正如彼得·施莱弗斯(P. Schryvers)对数字时代数据爆炸式增长后出现的数据陷阱的警示:“数字革命创造了一种信息狂妄。聚光灯越集中在这个世界可测量的部分上,我们就越相信我们无法测量的部分不再存在。”(施莱弗斯,2023:IX)在此,作者关注的是社会围绕数据驱动系统的组织方式,而非关于数据化、抽象化的道德问题。例如,对智能自我最关键的影响在于他人如何使用、为了谁的利益使用自己的数据。领域感知系统和持续监控系统可以实时跟踪城市中的任何人和任何地点,任何个体、家庭、组织、行为、事件的画像都可以通过数据驱动系统得到。这时,城市便被数字技术之网捕获,决定城市生存与发展的权力被掌握在数字技术和数据所有者的手中。

值得注意的是,作者并未谈及人的精神特性和社会关系性,未指出数据人和个体的本质不同,也未对人性进行具体分析。实际上,人是具有精神性的存在,精神和思想恰恰是无法通过数据精确衡量的变量。除此之外,现实的人是具体的、活生生的和处于各类社会关系下的存在,以数据来衡量个体强化了人对自我的机械认识,淹没了人之为人的思想特性和价值理性。人应是目的而非手段。这类人文情怀的呐喊和伸张,作者并未赋予笔墨,而是将其作为批判的理论前提

和情怀之基,以便重点关注资本主义政治经济体系的运行机制和现实。因此,从批判的现实性上看,作者并未过多游走在乌托邦式的遐思中。

整体而言,《过度智能》一书中对数字资本主义、互联网平台经济、数字系统的控制思想进行了深刻的分析和尖锐的批判。“新房东”和“分体”互为补充,成为数字资本主义批判的理论基石,并为理解数字劳动、数字泰勒主义、数据人等问题提供了有益的探索和启示。

三、技术批判的理路:技术、资本和政治

纵观人类历史,技术的发展与人类社会的进步紧密相关。每一次技术的巨大跃迁都塑造了新的社会管理模式、工作方式、生活方式和思维方式。对技术的反思、批判和完善一直是人类社会和人文学者关注的话题。因此,有必要梳理技术批判的理路,并将《过度智能》中对智能技术的批判与其进行对话,以期挖掘更深层次的内涵。

马克思在政治经济学框架下对技术进行了辩证论述:一方面,他认为技术作为物质生产的重要力量,是人获得自由必不可少的进步要素和手段;另一方面,他对技术进行了批判,并在《资本论》中对机器技术对工人的损害进行了揭露,认为“在工场手工业和手工业中,是工人利用工具,在工厂中,是工人服侍机器。在前一种场合,劳动资料的运动从工人出发,在后一种场合,则是工人跟随劳动资料的运动。在工场手工业中,工人是一个活机构的肢体。在工厂中,死机构独立于工人而存在,工人被当作活的附属物并入死机构”(马克思、恩格斯,1975:463)。技术的进步带来了生产方式的革新,进而造成社会关系的更新和变革。某种程度上,工人和技术之间的对立正是技术异化的必然结果。在劳动异化的进程中,技术不断被资本化。换言之,技术转变为资本扩张的动力和工具。技术的资本性利用和展开是造成劳动异化的原因之一,也是马克思对 19 世纪技术复杂性的揭示。

在 21 世纪的智能社会,技术的资本化是否在加速推进、变得更加深刻?作者声明,此书是关于政治、权力和利益的书,它们是技术的构成部分,并且也被技术所引导和改变(萨多夫斯基,2022:5)。毫无疑问,技术的资本化作为社会建

构的因素内含其中。不过,作者在这里使用了“利益”一词来解释技术的资本化,即技术实质上是实现利益的一种方式(萨多夫斯基,2022:6),比如少数人的利益、某些平台资本家的利益。因为技术并非自我决定的,而是由人们做出的各种选择的结果,所以人们行动的利益考虑、动机、原则等因素是否公平,维护的是哪方的利益便成为值得深究的问题。事实是,智能技术往往由少数人参与和制定,技术构建中存在着一定的寡头政治现象。这种现象往往被提供快捷服务的 APP 和网站所掩盖,算法、互联网带给用户的沉浸式体验淡化、弱化了人们对技术背后的政治性和利益性的关注。作者并未深究技术政治经济学的理论支撑,似乎是将技术的资本化作为资本主义政治经济体系默认的理论依据,关注的是技术政治化以及技术背后的利益受众、组织和价值等现实问题。

法兰克福学派继承和发展了马克思的社会批判传统,在 20 世纪 20—60 年代对技术“经历了一个从意识形态的分析与批判,到心理机制的分析与批判,再到对日常生活与大众文化的分析与批判过程”(王振林,2021)。例如,不同于马克思以商品为出发点,从交换价值分析人与物的关系,注重资本对生产逻辑控制的批判,鲍德里亚(J. Baudrillard)以符号编码为出发点,从文化角度、物的符号价值和象征意义分析消费行为和消费异化,展开了对消费的社会性批判。鲍德里亚指出:“消费是用某种编码及某种与此编码相适应的竞争性合作的、无意识的纪律来驯化他们;这不是通过取消便利,而相反是让他们进入游戏规则。这样,消费才能只身取代一切意识形态,并同时只身负担起使整个社会一体化的重任,就像原始社会的等级或宗教礼仪所做到的那样。”(鲍德里亚,2000:90)马尔库塞(H. Marcuse)在《单向度的人》一书中指明了发达工业社会的意识形态欺骗,认为“社会控制的现行形式在新的意义上是技术的形式”(马尔库塞,2008:9),而“这种技术秩序还包含着政治上和知识上的协调,这是一种可悲而又有前途的发展”(马尔库塞,2008:3),在技术主导的极权主义社会中产生着丧失否定性、批判性和超越性的单向度的人。在此,法兰克福学派默认了技术的建构性以及资本化,同时带来的新命题是:随着经济的发展,技术控制形式不断更迭,程度日益加剧。无论是消费社会的符号价值,还是虚假需要的制造,都是技术支配人们、操控社会秩序的手段和方式。

那么,21 世纪的智能社会是否日益被技术所控制?如果是的话,是借助什

么样的工具实现的？其形式呈现出什么样的新特点？作者认为，智能技术的快速发展日益控制了个体自我、家庭生活空间和城市空间这三大主体场域。例如，上文论述的分体和微观管理思想呈现了数据驱动的系统决定和改变了自我的评级、价值和认知，自我不知情地深陷其中且无法凭一己之力摆脱技术控制。智能技术通过数据最大化原则，尽可能地收集所有人、所有物的数据，实现了数据和算法驱动下的控制，结果便是技术所有者的权力更广和更集中。或许，智能社会中的微观管理在某种程度上实现了威胁、算法规训和大数据控制这三种权力的集合。

总而言之，从观照技术批判的理路中可以看到，作者认为智能技术一直起到了充实和赋能资本的作用，并通过数字平台、大数据技术实现了对个体、家庭和社会的微观控制和管理。更进一步，作者揭示了智能技术背后的政治和利益问题，即数字技术以提取所有人、所有物的数据并进行控制为要务，促进的是技术官僚和寡头的利益，损坏和凌驾了人类的民主、自治和公共利益。这为将政治纳入智能技术的构成部分，理解智能技术发展的政治性，提供了深刻和细致的批判范式。

四、批判之光与解构之行

马克思指出："理论一经掌握群众，也会变成物质力量。理论只要说服人，就能够掌握群众；而理论只要彻底，就能说服人。"（马克思、恩格斯，2012：9—10）能够切中要害、针砭时弊的理论，是引领人们进行精神革命的力量。《过度智能》一书是对当前以美国为代表的资本主义政治经济体系中有关政治制度、经济体制、民众生活、智能技术发展等多方面的综合性审视，运用大量案例将技术纳入社会的多元因素中予以批判，揭露了资本主义通往社会不平等、经济不稳定和生存不安全的黑暗面，并试图建构基于批判之上的新范式来为改变现实状况点燃亮光，提出了一些具有操作性和建设性的路径，一定程度上填补了理论与实践的沟壑，这也为处于数字化建设中的中国提供了有益的参考和借鉴。

一方面，技术创新民主化解决的是主体性问题，是从技术创新的主体参与设定上来实现技术为民所用。创新是少数人的专利吗？普通人也可以有知识和能

力做出创新吗？作者特意重申“卢卡斯计划”，意在提醒普通人拥有的智慧与力量，反对将创新神秘化和天才化。真正参与式的设计，必须从制度、监管、流程、能力等多方面来保障多元主体参与到技术的价值观、目标、生产方式等的设定和决策中去。另一方面，相较于数据监管这一常规性操作，作者更强调数据集体化的重要性。社会中每个成员的数据现已成为企业、经济、社会运行的基础要素之一，而数据的所有权却为个别企业所有，数据产品的私有化与生产数据的社会化正是数字资本主义的根本症结。作者鲜明指出：我们如何使数据治理变得更加民主？答案是去商品化和集体化，成立专门的数据管理机构，让数据出于公共利益而被集体拥有，真正实现造福于人的宗旨（萨多夫斯基，2022：226）。

技术是由社会所建构的，与一国的政治制度、经济体制、法律政策等密切相关，以大数据、互联网和通信技术为基础的智能技术更是人类社会自身的产物。换言之，审视技术也是在反观人类社会自身，批判技术的负面作用也是在反省人类社会存在的问题。批判的维度始终应当是人类和社会不可或缺的特性，正如马尔库塞在《单向度的人》中对单向度、封闭式思维及话语的批判，强调理解、反抗、挑战现存事物才能有改善它的空间，因为允许矛盾的存在才会保存张力的空间。智能技术正在飞速发展，其普遍应用的后果和影响深不可测且无法完全把握。《过度智能》一书坚守了社会批判理论的传统，对发展势如破竹的智能技术敲响了警钟。

批判之光照映在解构之行上，作者所提倡的方案对于中国来说也具有较大程度的借鉴意义。在中国式现代化进程中，“中国式现代化所强调的‘规模、共同、协调、和谐、发展’等内涵，与数字经济所蕴含的‘普惠、平等、协作、共享、开放’等特征具有内在一致性”（刘颖、黄朝椿、洪永淼等，2023）。在中国这片土地上，需要汇聚多元主体的智慧和力量，脚踏实地地推进数字经济和实体经济的深度融合。

综上所述，一方面，《过度智能》一书揭示了美国技术政治体系的权力不公平问题，以及数字资本主义驱动下智能技术凌驾于普通民众、服务于少数精英的异化之实；另一方面，借助大量案例分析，作者现实化、具体化了对智能技术的影响，为具体设计、制造、应用技术提供了批判思维路径。但是，在解构的现实之行上，智能技术数据集体化这一建议在操作上条件尚未成熟，如何采取更有效的措

施来使技术发挥造福于人的积极作用依然任重道远。

参考文献

鲍德里亚,2000,《消费社会》,刘成富、全志钢译,南京:南京大学出版社。

蓝江,2023,《数字焦土和剩余数据——数字资本主义社会中现代人的生存》,《求索》第1期。

刘颖、黄朝椿、洪永淼等,2023,《数字经济赋能中国式现代化建设》,《中国科学院院刊》第10期。

马尔库塞,2008,《单向度的人:发达工业社会意识形态研究》,刘继译,上海:上海译文出版社。

马克思、恩格斯,1975,《马克思恩格斯全集》第23卷,北京:人民出版社。

马克思、恩格斯,2012,《马克思恩格斯选集》第1卷,北京:人民出版社。

萨多夫斯基,2022,《过度智能》,徐琦译,北京:中译出版社。

施莱弗斯,2023,《为什么数据会说谎:被忽视的衡量指标》,张羿译,北京:中信出版集团。

王振林,2021,《一种时代的批判意识:当代西方马克思主义社会批判理论研究》,《吉林大学社会科学学报》第4期。

叶泽雄、鄢然,2023,《批判与超越:对数字资本主义的历史唯物主义检视》,《理论月刊》第5期。

通过人类行动思考机器智能

——专访人类学家露西·萨奇曼

露西·萨奇曼　胡万亨*

随着机器学习技术的快速发展和广泛应用，人工智能已成为当今人文社会科学领域的热点议题之一。实际上，早在20世纪70年代，以专家系统（expert system）为主流的早期人工智能研究就吸引了一批杰出的人文社科学者的关注，他们从哲学、社会学和人类学等角度做了深入的分析，对于我们理解人工智能在当下的发展仍有重要的借鉴意义。其中，著名人类学家露西·萨奇曼（L. Suchman）在80年代对美国施乐公司（Xerox）帕洛·阿尔托研究中心（Palo Alto Research Center，PARC）一款基于专家系统的交互式界面打印机的研发过程进行的详尽的民族志研究（Suchman，1987；Suchman，2006），是这方面的经典作品之一。萨奇曼目前为英国兰卡斯特大学社会学系荣休教授，近年来持续关注和研究人工智能技术的进展。她是人机交互（human-computer interaction，HCI）领域的重要开拓者之一，曾获美国语言学会“计算机与认知科学本杰明·富兰克林奖章”和美国计算机协会人机交互特别兴趣组“终身研究奖”。同时，萨奇曼长期活跃在科学技术学（science and technology studies，STS）领域，是科学的社会研究学会（Society for Social Studies of Science，4S）“贝尔纳奖”（Bernal Prize）的获得者——这是该学会的最高奖项，历届获奖者包括STS领域的奠基者以及为增进对科学技术之社会维度的理解做出杰出贡献的学者。在本次访谈中，萨奇曼分享了自己作为一名人类学家在施乐公司20余年的学术研究和工作经历，对包括专家系统和机器学习在内的人工智能技术的长期思考，围绕当代自主武器系统（autonomous weapon systems）的批判性研究和社会活动，以及对人文社科学者如何研究及影响技术实践等话题的看法。

* 露西·萨奇曼（L. Suchman），英国兰卡斯特大学社会学系。胡万亨，美国康奈尔大学科学技术学系。

一、科技行业的人类学家

胡万亨：您在加州大学伯克利分校期间，从本科到博士接受的都是人类学训练。您是如何开始对技术，尤其是对机器和人工智能产生兴趣的呢？

萨奇曼：这其实挺偶然的，可能对所有的博士生来说都是如此。因为我刚开始读博时，根本就没有考虑过技术。20 世纪 70 年代，美国人类学界掀起了一股叫“向上研究”（studying up）的浪潮。这个说法出自人类学家劳拉·纳德（L. Nader）70 年代发表在《重塑人类学》（*Reinventing Anthropology*）这本书中的一章（Nader, 1974），她是我在加州大学伯克利分校第一年的导师。这本书收录了一系列文章，以反思人类学领域，尤其是美国人类学的政治性。由于人类学有着殖民主义的历史渊源，这本书试图思考如何转变这一领域，使之服务于其他不同的政治价值。

纳德的论点是，人类学应该将注意力转向权力机构、权力中心。我深受启发。我那时也受到了互动论（interactionism）方法的很大影响，这跟当时人类学的主流取向有所龃龉。我在伯克利最后选了杰拉尔德·贝里曼（G. Berreman）做导师，他是一位互动论人类学家，从互动论视角对印度的种姓问题做了非常出色的研究。他对我非常支持。我当时有个想法，即用一个在我现在看来属于后结构主义的方法来看待权力的日常运作机制，更具体地说是企业中权力的运作。

不过，作为加州大学伯克利分校的学生和 20 世纪 60 年代成长起来的孩子，我跟任何大公司的联系都为零，我没有任何在大公司的人脉。但我发现，我其实还是可以通过学校的渠道联系到大公司的——伯克利分校人类学系的语言实验室跟施乐公司的帕洛·阿尔托研究中心有合作。于是，我们人类学系的博士生就可以通过这层关系进入 PARC，我之前的文章写到过这一点（Suchman, 2013）。

由于伯克利分校跟施乐公司的关系，我就直接去了 PARC。我对 PARC 可以说是一无所知，而且那时我连计算机都没有见过呢——在我那个年代，这可想而知，对吧？那个时候计算机还不普及。我来 PARC 根本就不是出于对技术的兴趣，只是后来才发现这里的技术真的太值得研究了。可以这么说，我在某种程

度上偏离了最初的目标。尽管我在那里学到了很多关于公司的知识,但我还是有些“分心”了,我被 PARC 当时在交互式计算(interactive computing)领域所进行的工作分散了很多注意力。

那个时候正是交互式计算发展的早期阶段。1978 年我初到 PARC 时,人机交互的想法还非常新,我立马就对这种将计算机界面作为互动场所的理念产生了兴趣。我开始觉得,对人与计算机接触的过程中发生的事情进行一些互动论分析,应该会非常有趣。与此同时,我所在的实验室里也有认知科学和人工智能方面的研究项目。那时还是规划模型(planning model)在人工智能和专家系统中占主导地位的时代。

慢慢地,我越来越多地参与到这个项目中来。它要打造一个能够指导用户使用人机合作型复印机的专家系统原型,这在当时听起来还是挺疯狂的。就这样,我对研究技术产生了浓厚的兴趣。与此同时,我接触到了 STS,它给我的研究建立了一个参考框架,也让我加入了一个由其他社会科学家组成的圈子,他们以非常有趣的方式思考技术。

胡万亨:您最初感兴趣的是用互动论视角去考察权力运作,为了研究企业中的权力而来到施乐公司。当您看到人类与机器或复印机之间的互动时,发现用互动论来思考也很合适,这样说起来确实很自然。

萨奇曼:没错,我立马就意识到了这一点。并且,我对人机互动过程中所发生的事情,在多大程度上类似于人人互动是持怀疑态度的。这就是我的研究的起点。这个“类似”在哪些方面可以成立,它的限度在什么地方,又在哪里无法成立?这就是我的博士论文要回答的问题。

胡万亨:您对人机交互产生兴趣时,还在一个人类学项目里学习。在 20 世纪 80 年代,作为人类学家去研究技术是种什么感觉?我的印象是,这在当时应该不是传统的路数。人们对这类研究的反应如何?以及,那时的技术人类学领域大致是个什么情况?

萨奇曼:确实不太传统。当然,那时候已经有一些重要的学者做了相关研究。比如,莎伦·特拉维克(S. Traweek)基于对一个粒子物理实验室的民族志

研究,出版了《物理与人理:对高能物理学家社区的人类学考察》(Traweek,1988)。但这本书更多是对科学实践的民族志,它的田野地点虽然不寻常,但就方法而言,我觉得它比我的研究更为传统。幸运的是,我的导师很有同理心,他的思想非常开放,对我也很支持。尽管我的博士论文离他所做的研究越来越远,我们仍然有互动论的共同研究取向,而且我们在关心的事情上有着相近的政治立场。他调解了我与系里的关系,帮助我渡过了博士生所要经历的各种关卡。但不得不说,与 STS 的联系对我来说越来越重要——那时不断发展的 STS 圈子,在研究取向上跟我正在研究的东西要一致得多。

还有一点,我的博士论文委员会里还包括赫伯特·布鲁默(H. Blumer)——一位出色的符号互动论学者,以及哲学系的休伯特·德雷福斯(H. Dreyfus)。所以,我的博士论文委员会里有着更了解我论文的实质主题的老师——不用说,尤其是德雷福斯。

胡万亨:原来如此!我们来谈谈您在施乐公司实习的经历吧!您在做参与观察和民族志研究时,肯定也会有一些具体的工作职责吧。您在施乐公司具体做什么呢?

萨奇曼:这是一个非常有趣的过程,也是我融入 PARC 这个环境的过程。跟我找到这份实习工作一样,我的工作内容也是有迹可循的。团队成员之前也接触过人类学家,因为在我来之前,埃莉诺·韦恩(E. Wynn)就曾在 PARC 待过,她的博士论文是关于办公信息系统和办公自动化的。我最开始实习时,加入的是“办公室研究小组”,那个时候还是办公自动化的早期阶段。我的同事对工作实践(work practices)很感兴趣,这在某种程度上是一种基础性的研究工作。但我到 PARC 后不久,“办公室研究小组”就解散了——企业界就是这样,上头下来个什么指令,一切就风云突变了。所以,我那时基本上是被另一个小组收留了,也就是约翰·布朗(J. Brown)领导的“认知和指导科学小组”。

我刚到时,基本上就是凭感觉干活。那是个暑期实习,我是埃莉诺的研究助理。她刚刚完成博士学位论文,正在做一项研究——施乐公司那时正在研发 Alto 计算机,并将其视为办公信息系统。他们决定进行一系列所谓的“探索”(probes),也就是将 Alto 引入工作场所。一开始,这些基本上都是在施乐的行

政办公室进行的,属于某种内部探索或实验。埃莉诺在施乐位于湾区的一个行政销售办公室做了一系列预访谈,然后给他们装备上 Alto 计算机。一段时间后她再回访,看看会发生什么。我最初是作为暑期实习生来转录这些访谈的,我在一台 Alto 上做了转录,用的是一个名为 Bravo 的早期文字处理包。

这份转录工作是一个学习实际使用 Alto 的机会——它还蛮好用的——也是一个在暑假期间结识同事的机会。在那段时间,我一直在留心,想弄清楚我自己在那里实际上可以做什么。暑期实习结束后,我获得了一个研究实习的机会,来进行我的博士学位论文研究。

所以,我经历了几次不同的转变。从一开始更多关注工作和自动化,到后来更多关注人机交互、人工智能和认知科学领域。PARC 是一个非常适合探索和思考的地方:一方面,我有很大的自由去关注有趣的事情;但另一方面,我必须不断努力让同事能够理解我,并且思考如何把我关心的问题变成一个博士论文题目,让我的博士论文委员会也可以理解,觉得有趣,同时还要结合 PARC 这边正在进行的实际工作。也就因此,这是一个共同打造研究项目并找到进入田野空间的方法的过程。我很幸运能够真正被施乐雇用,也同时攻读博士学位,很幸运能够来到这个非凡的地方,拥有如此多十分有趣的可能性。

话说回来,你也得时刻准备着,找到一些方法让自己变得容易理解,让周围的人对你的工作感兴趣。这些人基本上都是计算机科学家,还有一些是认知心理学家。所以,你得做大量的“转译”工作,而且尽量不要单纯地去迎合他们的想法,把自己完全“转译”成他们眼中“我可能是谁”“我可能在做什么”的样子。

胡万亨:您毕业后继续留在施乐公司工作,并且度过了 20 年的职业生涯。您能谈谈决定留在科技行业尤其是施乐的考虑吗?

萨奇曼:当然。我完成博士学位论文后,拿到了施乐的一个实际上相当于博士后的职位,得以暂时延续我在那里的工作。那时我对施乐已经足够熟悉,而且施乐也有足够多的潜在研究方向,我也有能力推进这些方向。所以,继续留下来就是一个自然而然的选择,这是一个继续推进我手头工作的绝佳机会。

我的博士后差不多快结束时,我跟当时的上司约翰·布朗说:“我在想博士后做完后的安排。”他大意说:“哦,我们给你提供了一个长期职位。我还没跟你

说吗?”所以,他其实已经直接把我调到了长期岗位上。这个情况之所以发生,是因为当时 PARC 还处于早期发展阶段,还没有以任何方式规范他们的招聘流程。我觉得当时应该没有人反对,但这个决定确实是在十分非正式的基础上进行的。那时,我已经深深地融入了公司,并且可以看到很多有趣的方向,所以完全没有问题。我对能够待在那里非常非常开心。正如你提到的,我在那里度过了接下来的 20 年。因此,从我拿到博士学位过渡到人生中的下一个重要阶段,顺利得出乎意料。

胡万亨:您在 2000 年加入兰卡斯特大学,回到了学术界。您是如何决定重返学术界的呢?

萨奇曼:我会说有推力——有拉力,也有推力。对我和我的社会科学同事来说,维持与 PARC 和施乐公司的关系变得越来越困难,我们越来越多地被要求扮演服务性的角色。就像我之前说的,这类角色对今天工业界的社会科学家来说是最常见的。彼时,我们在 PARC 成立了一个名为“工作实践与技术”的研究小组,旨在探索如何与相关工作场所中的用户合作开发一些信息技术原型机。

在此期间,我们与施乐最感兴趣的一些组织开展了一系列项目,与施乐产品项目部门以及全公司出色的界面设计师密切合作。在这样的合作过程中,我们小组成员之间变得密不可分。我们有一些自己的长期项目,而且我们之间的紧密联系也是非常必要的,因为我们想要跟加州交通部的土木工程师或硅谷一家大型律师事务所的律师建立合作,我们决心开展几个长期项目。与此同时,PARC 的风向也在发生变化。PARC 的研究议程经常变换,这会儿说“每个人都应该去那里”,过会儿又说“每个人都应该朝另一个方向去”。我们对此持抵制态度,我们坚持遵循此前积累下来并且认可的研究程序。

就我个人而言,一个决定性时刻是我跟当时的经理交流意见时,他告诉我,他认为我抗拒改变!他怎么能说我抗拒改变呢?但我意识到,当他这么说时,他的意思是我抗拒改变以适应新的议程和新的工作方式。我想,我和我的研究小组那时都感到,20 世纪 90 年代末的 PARC 的发展方向跟我们越来越背道而驰了。

最后,我们集体辞职了。我们也想过要一起去做一些事情,因为我们喜欢一

起工作。不过,失去了 PARC 这个平台,我们终究还是各奔东西了。就我个人而言,我一直很清楚,如果离开 PARC 的话,我还是想回到学术界。所以,在 PARC 工作期间,我也非常努力地维系回归学术界的桥梁。我之所以知道兰卡斯特大学,是因为他们社会学系的一些人跟“计算机支持的协同工作”(Computer Supported Cooperative Work, CSCW)会议联系密切。他们做了一些非常出色的工作,比如约翰·休斯(J. Hughes)等人对空中交通管制员的技术变迁的研究。

兰卡斯特大学还有一个非常强的 STS 项目,有约翰·劳(J. Law)与许多其他鼓舞人心的 STS 研究者,以及性别和女性研究方面的杰出同事。因此,当兰卡斯特大学发布了两个职位的招聘公告时,我就想我肯定会申请。这就是我最终来到兰卡斯特大学社会学系的原因。这是一个非常棒的系,而且那些真正对跨学科思考感兴趣的人遍布整个大学。我立即感到宾至如归。与此同时,我从周围的每个人身上都学到了很多东西。所以我在那里度过了接下来的 20 年!

二、人类行动和机器智能

胡万亨:现在我们来谈谈您的具体研究吧!先聊聊您的经典作品《人机重构》(*Human-Machine Reconfigurations*)——最早的版本标题叫《规划和情境化行动》(*Plans and Situated Actions*)。书中对用“规划模型”(planning model)去理解人类行动,由此开发智能机器的想法提出了批评。您能简单解释一下书中的核心论点吗?

萨奇曼:规划模型的基本论点跟有关交互性(interactivity)的看法密切相关。当时在人工智能领域占主导地位的就是规划模型,它对人类智能和相互可理解性的概念有着非常强烈的认知主义色彩。它的基本想法是,所有真正要紧的事情都发生在我们的头脑中,在我们的大脑或心灵中,人的智能行动是目标导向的,是按照制定好的规划有效执行的。行动的要点就在这些规划之中,随后被执行出来。这种认知主义的观点能很好地映射到计算机技术的能力范围内,因为规划可以对应到程序里,计算机程序是对行动要求的具体化。人们也认识到,在规划执行过程中可能会出现障碍,可能会遇到必须适应的环境条件,但这些都是

行动者执行计划时需要解决的问题。具体到沟通行为上,它意味着我们相互理解的方式是通过认识彼此的规划。

我当时参与研发的系统就是基于规划模型的。回到复印机的例子,当我们进行复印时,我们有一个规划,有一个目标,有一种我们想要执行的具体规范。我们制定了实现这一目标的规划。因此,我们就想,专家系统需要能够识别用户的规划,然后为他们提供必要的指令,以使用这台机器执行该规划。我的分析就是基于人们在实际理解操作界面并试图使其变得易于理解时遇到的困难而展开的。

我当时对常人方法学和互动论分析很感兴趣,所以我录制了一些人们尝试使用复印机的视频来研究这个问题——最开始是比较原始的复印机,后来的复印机加入了专家系统。让我不可思议的是,我坐下来观看这些录像带,而且如果我足够仔细地观察,我就能弄清楚这些人想要做什么。然后,我发现他们实际上并没有按照设计者的意图来理解机器,机器的很多"举动"用户也并不理解。所以,我可以看到所有这些问题。我开始思考,我为什么能看到所有这些麻烦,但机器似乎完全没有注意到,专家系统似乎也完全没有,或者错误地识别了用户的行为。

我开始更仔细地观察这些互动,我在书中用一个四列表格列出了这些观察。我录制的视频里,人们要么在谈论他们正在做的事情,要么是两个人试图一起做点什么,所以在互相交谈。我可以听到他们彼此交谈的所有事情,看到他们围绕机器所做的所有事情,我将它们转录到表格的第一列中。然后,我可以看到其中那一小部分机器可以在现实中做到的事情,我将其转录到第二列中。当然,因为它只知道是什么(指令)会改变它的状态——即便是我们现在使用的更加复杂的计算界面,也还是如此。在我们人的世界中,唯一能跟机器交界的,就是那些能够实际影响机器状态的方面,比方说敲击键盘、点击触摸板或是其他之类的操作。所以,我开始认识到,这个系统只有一个小得令人难以置信的小孔,这是它能够了解正在发生的事情的唯一通道。

接着,我在表格的第三列指出了用户哪些操作得到了机器系统的响应。最后,在第四列,我列出了这些操作背后的设计原理,这些原理都是基于把用户的操作理解为对他们头脑中规划的执行,进而给系统设计相应的响应。如此,我能

够看到用户和系统之间的所有不一致(misalignment)。我由此得出的基本论点聚焦在人与机器之间存在的极大的不对称性(asymmetry)上:当我们参与行动、理解彼此以及世界时,我们有很多可以使用的资源,而计算机则少之又少。

基于对这种不对称性的观察,我对于规划与情境化行动的论点在很大程度上得益于哈罗德·加芬克尔(H. Garfinkel)的常人方法学中一个非常基本的观点:任何关于行动的规定——无论是说明、法律、规则还是规范——都必然以找到规定与当前具体情况之间关系的能力为前提。说明书上的指示预设了用户的这种能力。这些说明没有,也永远不可能,完备地具体指明如何找到这种关系。因为,不论是遵照什么样的指令,用户面对的都是一个不可简化的开放世界。自此之后,我所有的研究都贯穿着这个前提,即计算系统在某些重要方面需要在封闭的世界里运行,而计算机与用户交互的界面,本质上就是对开放世界、对我们在世界上的行为和活动的转译,使之能被计算机读取。

胡万亨:说到这本书本身,我很喜欢您在书中开篇举的例子,您谈到欧洲人和特鲁克人(the Trukese)如何以不同的方式在海上航行。

萨奇曼:是的。这是人类学里的一个经典,托马斯·格拉德温(T. Gladwin)和其他人描述了南太平洋的航海家如何在没有地图和仪器的情况下,不可思议地从一个岛屿准确到达另一个岛屿。

简单来说,这个故事说的是他们的参照点是周围的所有事物,如海浪、星星、风、太阳等等。这个例子的要点并不是说世界上有两种人,一种人总是按规划行动,另一种人总是看情况行动;相反,这个例子比喻的是我们(作为西欧人)思考目标导向之行动的方式,以及思考它的不同方式。在一些重要的方面,我想,我们所有的行动都涉及调动不同类型的资源,根据具体的情况去寻找资源。这始终是一个将我们可用的工具应用于所处的实际情况的问题。

胡万亨:其他一些学者也对人工智能提出了一些著名的批评,如您所知,比如德雷福斯、哈里·柯林斯(H. Collins)和戴安娜·福赛斯(D. Forsythe)。相对于他们的看法,您如何定位自己的观点?

萨奇曼:我当时也读到了他们的作品,深受他们的影响,从中汲取了很多灵

感,尤其是从德雷福斯那里。我想,我们可以先从他的《计算机不能做什么》(Dreyfus, 1978)及其续集《计算机仍然不能做什么》(Dreyfus, 1992)说起。他带来的现象学视角,真正阐明了认知科学和人工智能是如何消除具身性和情境性的。我认为,他以一种非常有力的方式指出了计算的局限性,他的论点到今天仍然站得住。随后,他对联结主义(connectionism)的出现给他的论点带来的影响和挑战感到相当不安。我认为,他对此多虑了,他的论点仍然是成立的。

然后是柯林斯和福赛斯。我们每个人的出发点都略有不同,但殊途同归。柯林斯是一位知识社会学家,带来了知识社会学的视角(Collins, 1990)。在研究人工智能之前,他的《改变秩序》(Collins, 1985)一书是关于科学实验的复现的,但同样也是关于具身性和情境性的。那本书对我来说真的很有启发。福赛斯出身于更传统的文化人类学领域,她对当时的人工智能有很深的研究(Forsythe, 2001)——那时候还以知识工程(knowledge engineering)和专家系统为主。因此,她把对知识构成的人类学理解带了进来。这同样可以追溯到德雷福斯那里,并且与我提出的论点也非常有共鸣。

所以,我们提到的这些书,无论是哪一本,我读起来都觉得非常有共鸣,有很多共通之处。我们对人工智能的批评只是在取向上略有不同,主要是我们所借鉴的知识传统不一样。让我受益匪浅的另一本书是艾莉森·亚当(A. Adam)的《人工认知:性别与思维机器》(Adam, 1998),这本书非常精彩。艾莉森是一名计算机科学家、软件工程师,具有非常强烈的女性主义意识——她阅读了大量女性主义学术著作。

胡万亨:德雷福斯在他的《计算机仍然不能做什么》一书中谈到了机器学习。柯林斯于2019年出版了他第二本关于人工智能的书《人工虚拟智能》(Collins, 2018),主要讨论的就是机器学习。遗憾的是,福赛斯似乎并没有对此说太多,她去世得太早了。我想知道:您如何评价机器学习?如果您今天要再写一本关于人工智能的书,您之前的论点如何适用或不适用于机器学习?我想听听您对这个领域的现状有何看法。

萨奇曼:这是一个非常有趣的问题!因为从专家系统、知识工程和符号主义人工智能时代,到现在这种数据密集型、高算力模式的人工智能的爆炸式发展,

很多方面的情况都已经发生了翻天覆地的变化。但我的看法还是一以贯之的，还是可以追溯到封闭世界和开放世界的问题。

如果我们看看大语言模型，看看它的自我参照性（self-referentiality），不同之处也就在于，现在这些机器学习系统的能力不是依赖基于规则的操作，而是可用数据的规模以及能够从中识别相关性的计算统计，然后进行参数化和迭代以产生可理解的输出。这就是我对它的核心特征的描述。

我们仍然可以将这些模型视为一种封闭的世界系统、一种自我参照的系统。现在这些模型所谓的“脆性”（brittleness），就是在它们遇到自我参照性的局限时显现出来的。事实上，大语言模型基本上是在以数字方式捕获的文本和图像上运行的，而完全不是在它们所指涉的真实世界中运行的。

因此，正如我此前研究的专家系统只能依照规划模型读取指令从而在不同的状态间切换而实现运作一样，当前的系统只能根据它们运行的数据集和数据模型中存在的相关性进行响应。它们其实什么都不知道，对吧？我的意思是，它们对所有这些数据所指涉的实质内容其实一无所知。有趣的是，现在已经有一些人在呼吁回归到混合模型上来，比如加里·马库斯（G. Marcus）这样的关键人物（Marcus, 2018）。数据驱动的人工智能之路正在遇到瓶颈，在寻找下一步对策时，该领域的人们现在正回头转向符号人工智能——“好吧，现在我们需要回到逻辑”——不论如何，它仍然还是会处在一个逻辑关系或统计相关性的封闭世界中。

我觉得观察这一切、追踪这些巨大的变化真的非常有趣——这些变化同时也是我们如何生活和穿梭在开放世界与封闭世界中的方式的变迁——当然，我认为有些东西也没怎么变，有一些非常根本的方面并没有改变。还有一点没有变的是，人工智能界还是不愿意承认这一点，他们对类人智能的渴望和决心如此强烈。不知何故，在他们看来，如果承认这些局限性，就会贬损这项奇妙的技术——我认为它真的非常了不起。这些局限性在我看来，只不过是计算系统和作为有机体的动物之间永远都会存在的差异而已——我说的“动物”不仅仅是指人类，还包括人类之外的动物，甚至其他更多非人类世界的存在。

三、跨越学科边界，连接学术与社会

胡万亨：我们讨论了您早期对符号主义人工智能的批评，也探讨了今天的联结主义人工智能。当时的工程师如何回应您的批评呢？这对当今的人工智能研究者有怎样的启发？

萨奇曼：我认为“因人而异”总是非常重要的——人们会问：施乐对这项工作有何看法，或者 PARC 有何看法？但实际上，我们说的都是特定的一些人。你会遇到一些非常关键的搞技术的人——现在回想起来，我尤其感谢菲尔·阿格雷（P. Agre），他在我的思考过程中非常重要。我们都还在念博士的时候就认识了，我从他那里得到了很多灵感，也学到了很多东西，他那时正在麻省理工学院攻读人工智能方面的博士学位，而我在加州大学伯克利分校攻读人类学博士学位。在像麻省理工学院或 PARC 这样的地方，有计算机科学家，有非常投入的工程师，但他们也会有疑问，对吧？他们也会觉得在工作中找到问题和改进的机会，以及思考自己工作的新方式更有趣、更令人兴奋。

我在 PARC 的任务主要就是找到这些人。一路走来，我发现他们中的许多人确实对重新界定他们正在解决的问题或以不同的方式思考解决方案很感兴趣。我得说，那些人真的就是我的盟友。即便是理查德·菲克斯（R. Fikes）也对我的批评非常感兴趣，而他是开发我所研究的专家系统的、秉持人工智能的规划模型的人，他也是一个思想非常开放的人。

所以我认为，与其将人工智能社区视为一个同质的整体，不如与那些有兴趣批判性地反思自己实践的技术从业者建立联系。而且，他们也可以给你讲这些工作的进展，以便你能够充分了解田野的情况并提高自己的技术素养。刚去 PARC 的时候，我甚至觉得自己得成为一名会 LISP 语言的程序员，如此才能真正了解这里发生的事情。

后来，我意识到不用成为一名 LISP 程序员，因为我自己就是一个人类学家而不是计算机科学家啊！我当时觉得——也许有点事后诸葛亮了——哪怕我自己对这些技术只掌握了一点皮毛也没有关系，因为我身边都是世界上最好的 LISP 程序员，对吧？因此，我要做的就是去合作。于是我有了像兰迪·特里格

(R. Trigg)这样的合作者,我和他一起工作了很多年。我们共同撰写了“作为手艺的人工智能”(“AI as Craftwork”)一章(Suchman & Trigg, 1993)。兰迪是一个可以和我一起思考的人,他有这个知识和能力。我觉得这种合作令人兴奋且富有成果。不得不说,司马贺(H. Simon)实际上也觉得这种合作是必要的,这也许是我成功的标志之一——他和艾尔松索·维拉(A. Vera)写了一篇题为《情境化行动:一种符号主义的方法》的论文(Vera & Simon, 1993)。他们的基本观点是,我和其他人所提出的关于情境化行动的论点可以毫无问题地整合到符号主义人工智能里并为之所用。我认为,这是一种防御性的举措,好比说我们必须得“中和”掉这种批评一样。有些人还是会坚持己见,但会试图接受任何批评并使其中性化,这样它就不会以任何方式干扰或破坏他们的进展。这是一种跟搞技术的人打交道的模式。但对我来说,更具创造性的模式是与像菲尔和兰迪这样的人合作,他们对自己的领域有着自己的反思,有兴趣以各种各样的方式重新思考它,而且是跟我们这样的人一道。

胡万亨:我很喜欢您关于“中和”批评的说法。如今,科技公司似乎越来越流行聘请一些社会科学家,让他们做用户研究、界面体验等等。您如何看待人类学家、社会学家或其他社会科学家在科技行业中的角色?您在施乐的情况在 20 世纪 80 年代很常见吗?

萨奇曼:我想说,我的例子以及施乐少数那几个像我一样隶属于 PARC 的研究人员的情况是相当特殊的。虽然 PARC 是一个由施乐资助的研究中心,但它的运作非常独立。那个时期施乐公司有大量资金,而当我 2000 年离开 PARC 时,情况完全不是这样——施乐公司几乎破产了。但早在 1970 年,PARC 刚成立时,有大量资金用于未来投资。PARC 非常独立,那里的每个人都有博士学位,简直就是个学术研究组织的延伸。我们都得在学术上有所成就,所以我们需要发表论文、参加会议,这意味着我们有很多机会保持与学术界的联系。这原本就是一个不同寻常的情况,在此后的 20 年间也确实改变了。当我离职时,PARC 更像是一家大公司内部的典型研发组织,非常营利导向。

我在 PARC 的职业生涯末期,公司希望它增加专利和许可证以及其他可以让研究变现的方式。正如之前所说,我加入时,它更像是一个学术研究机构。我

认为社会科学家融入工业界——我特地使用“融入”(incorporation)这个词——在很大程度上是营销方式的一种延伸,尤其是在设计方面,是“用户体验”行业成长的一部分。这个行业为拥有社会科学学位的人提供了就业机会。

一方面,我认为这很好,因为人们确实需要工作,社会科学给工业界和设计领域带来的意识观念很重要;但另一方面,我认为对社会科学家的投资比媒体所暗示的要少得多,那些被聘用的社会科学家是用户体验、界面设计业务的一部分,与工业界的其他活动一样,这些业务的时间框架非常短。所以,研究人员鲜有机会真正长期参与田野点的活动,也很少能生产出进入学术圈的知识,无论是人类学还是社会学。

因此,可以说,这样一些工作几乎没法给社会科学家提供充分的施展机会。我认为这并非幸事,但这就是业内的普遍状况,尤其是在公司利润增长放缓的情况下。做研究的空间变得越来越小,社会科学家通常都是第一批被裁的人,而且越来越多地需要为产品计划、品牌和营销提供有用的故事。

胡万亨:我们来聊聊您最近对自主武器系统的研究吧。您能简单介绍一下您在这个方向的工作吗?您正在解决的关键问题是什么?是什么原因促使您关注这个主题的呢?

萨奇曼:之所以做这个题目,可能是因为我在20世纪60年代的美国长大,是反战运动的孩子。但即使在80年代初的施乐公司,也有一些人在PARC从事这方面的研究。其中最著名的是计算机科学实验室的塞韦罗·奥恩斯坦(S. Ornstein),他曾从事过地面环境下半自动化武器系统(semi-automatic ground environment, SAGE)的研发,这是美国第一个计算机化导弹防御系统。80年代初,塞韦罗越来越担心所谓的“战略性防御计划”,该计划将计算机控制引入了核武器系统。

因此,我与塞韦罗以及PARC和斯坦福大学的其他同道一起参与创立了计算机专业人员社会责任联盟(Computer Professionals for Social Responsibility, CPSR)。当年有一些计算机科学家讨论了用所谓的“预警发射”来控制核弹的自动响应系统的诸多不可靠之处。多年来我一直积极参与CPSR,这些工作演变成了一系列追踪武器指挥和控制系统的自动化,特别是计算机化的项目,自主

武器系统是相关工作在当下的题中应有之义。对我来说,这是对自动化——尤其是自动化瞄准——以及越来越多的远程控制武器系统的出现的更广泛担忧的一部分,它有可能导致自动瞄准能力被一发不可收拾地用到武器系统中来。

我最近加入了机器人军备控制国际委员会(International Committee for Robot Arms Control)。这是一个由计算机科学家、物理学家和社会科学家组成的很棒的团体,其试图推动围绕自主武器系统建立具有法律约束力的条约——自主武器系统在目标识别瞄准和启动发射等关键功能上进行了完全自动化。

我们所呼吁的禁令,针对的是那些在关键功能上去除了所谓“有意义的人类控制”的武器。我深信这项运动的意义,它真的非常重要。与此同时,我对“战争背景下的有意义的人类控制”这个想法也有一些疑问。因此,我的研究往大的方面来说,涉及分析有哪些各式各样的投资、话语和技术在支持和维系“现代战争可以是一种受控制的理性事业”这一理念。我认为事实并非如此。

在这项工作中,我主要参考和对话的是保罗·爱德华兹(P. Edwards)的《封闭的世界》(Edwards, 1996),还有约瑟夫·马斯科(J. Masco)及其他 STS 学者的工作。这些研究追溯了这种控制战争的想法在二战期间和战后,以及从冷战到如今的发展。

我最近在写一篇文章,是关于军队中封闭世界思维的当代化身的。用我们 STS 的话来说,就是对一个完全仪器化、军队掌握了完备信息的世界的“想象”(imaginary)(Suchman, 2020; Suchman, 2023),我认为这仍然是一种幻想。我认为追溯这种思维的逻辑以及它为对新武器系统的投资、政策制定和国际关系提供理由,并揭示这种思维方式的谬误之处都非常重要。这种思想确实是站不住脚的。这些谬误之处为我们开辟了空间,去思考一些替代方案来阻止军事主义以及处理国际关系。

胡万亨:我们来谈谈一些更广泛的问题吧!您的研究对 STS 和 HCI 等跨学科领域做出了重要贡献,也获得了美国计算机协会和 4S 奖项的认可。在您看来,定义 STS 和 HCI 领域的核心问题分别是什么,它们有何不同?您是如何参与这些领域的对话的呢?

萨奇曼:这两个领域对我来说都非常重要。HCI 领域的大会 CHI(computer

human interaction)规模庞大,而且非常技术中心主义(techno-centric)。但是,CHI 内部有越来越多的人在致力于把人机交互问题放到一些更广阔的问题中去研究,比如系统开发的条件、用户是谁、用户所处的情境以及用户是如何被设定的。

HCI 领域有很多非常有趣、更富批判性的工作从 STS 中汲取了灵感。而且,跟以往相比,我觉得越来越多的 STS 学者也在寻求参加 HCI 领域的会议,以便更接近那些实际参与研发工作的人员。同样,从事设计的人员也在从 STS 中寻找资源,以不同的方式去思考他们正在做的事情,扩展他们的工作框架。在 PARC 期间,我主要活跃在 HCI 和 CSCW 领域。但我一直坚持把人类学和 STS 作为我智识上的灵感来源,所以我在这两个世界之间来回穿梭。这确实让我收获颇丰。如果没有从 STS 获得新的思考来源,我想我不大可能在 HCI 领域做出这些贡献。自从我离开 PARC 以来,我更多地活跃在 STS 学界。

所以,如今我已经不太参加 CHI 或者 CSCW 会议了。但随着信息学院的兴起,比如康奈尔大学的计算机与信息学院,现在已经有越来越多的交叉对话和研究。信息学院对于那些没法完全归入某个学科类别、处在类别边界的人来说,真的是非常理想的归宿。同时,这类学者在那些以往技术性更强的会议中的影响力也越来越大。

胡万亨:作为一名人文社科学者,要研究计算机科学和智能技术的话,您觉得最重要的技能、资质是什么,或者应该有哪些准备呢?

萨奇曼:我会说是谦虚。我的意思是,你得意识到自己在其他相关领域的知识掌握程度和局限性,并始终努力加深你的理解。同样重要的是,要学会寻找合作者。在 STS 领域里,有的学者技术素养非常高,对研究主题的技术细节有着深刻的理解。但我认为,很多情况下,我们中的许多人无法在短短一生中在技术领域掌握跟社会科学相关领域同等的能力。因此,我们要不断努力去加深理解,同时寻求富有成效的合作。我现在非常开心能与一些像你这样的年轻学者进行持续的对话。跟你一样,他们也身处跨学科的交叉领域,也在真正深入机器学习的技术细节中,同时充分保持和发挥 STS 的敏锐意识。这样就能够把 STS 中的一些思想概念引进来,从而强有力地、有创造性地去思考技术实践。

所以,我认为持续学习和建立联系可能是最重要的事情,也是你能拥有的最重要的技能。

胡万亨:的确,这里的关键词是“协作”。您对年轻学者有什么特别的建议吗?

萨奇曼:寻找志同道合的学术同辈,不是吗?这太重要了。如今的就业市场真的非常非常难。从我早先对自己在工业界工作经历的评论,你可以看到,我其实还是在学术界感觉更亲切。但我认为在工业界工作是一种非常有价值的经历,最起码也是宝贵的民族志。所以我想说,如果你确实很在意维持在学术方面的发展,你可以想办法在业界工作的同时跟学术界保持联络。

去参加 STS 会议并且报告论文,这就是我保持学术联系的方式,我就是确保自己能够出席并参与我想保持联系的学术活动。所以,我觉得,还是有很多办法可以找到工作、找到雇主,并保持你所关心的学术联系的。这确实有点挑战性,但也是可行的。

参考文献

Adam, A. 1998, *Artificial Knowing: Gender and the Thinking Machine*, London: Routledge.

Collins, H. 1985, *Changing Order: Replication and Induction in Scientific Practice*, London: Sage Publications.

Collins, H. 1990, *Artificial Experts: Social Knowledge and Intelligent Machines*, Cambridge, MA.: MIT press.

Collins, H. 2018, *Artifictional Intelligence: Against Humanity's Surrender to Computers*, Cambridge: Polity.

Dreyfus, H. 1978, *What Computers Can't Do: The Limits of Artificial Intelligence*, New York: HarperCollins.

Dreyfus, H. 1992, *What Computers Still Can't Do: A Critique of Artificial Reason*, Cambridge, MA.: MIT Press.

Edwards, P. 1996, *The Closed World: Computers and the Politics of Discourse in Cold War America*, Cambridge, MA.: MIT Press.

Forsythe, D. 2001, *Studying Those Who Study Us: An Anthropologist in the World of Artificial*

Intelligence, Stanford: Stanford University Press.

Marcus, G. 2018, *Deep Learning: A Critical Appraisal*, https://doi. org/10. 48550/arXiv. 1801. 00631.

Nader, L. 1974, "Up the Anthropologist: Perspectives Gained from Studying up." in D. Hymes ed., *Reinventing Anthropology*, New York: Vintage Books.

Suchman, L. &R. Trigg 1993, "Artificial Intelligence as Craftwork." in S. Chaiklin &J. Lave eds., *Understanding Practice*, New York: Cambridge University Press.

Suchman, L. 1987, *Plans and Situated Actions: The Problem of Human-Machine Communication*, Cambridge: Cambridge University Press.

Suchman, L. 2006, *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge: Cambridge University Press.

Suchman, L. 2013, "Consuming Anthropology." in A. Barry &G. Born eds., *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences*, London: Routledge.

Suchman, L. 2020, "Algorithmic Warfare and the Reinvention of Accuracy." *Critical Studies on Security* 8(2).

Suchman, L. 2023, "Imaginaries of Omniscience: Automating Intelligence in the US Department of Defense." *Social Studies of Science* 53(5).

Traweek, S. 1988, *Beamtimes and Lifetimes: The World of High Energy Physics*, Cambridge: Harvard University Press.

Vera, A. &H. Simon 1993, "Situated Action: A Symbolic Interpretation." *Cognitive Science* 17(1).