

作为基础资源的大数据： AIGC 变革下新闻传播活动的再认识

涂凌波¹, 赵奥博²

(1. 中国传媒大学电视学院, 中国传媒大学广播电视学系, 中国传媒大学媒体融合与传播国家重点实验室, 北京 100024; 2. 中国传媒大学电视学院, 北京 100024)

摘要: 作为自然语言处理技术和生成式人工智能的里程碑, Chat GPT 技术和 AIGC 的兴起正在改变新闻传播活动的面貌, 也意味着大数据已深度嵌入社会系统之中。文章从新闻传播学角度, 讨论大数据的角色和功能及其对新闻传播活动所产生的影响。在厘清大数据的四种基本认识基础上, 文章认为大数据是驱动数字新闻传播活动的基础资源和关键要素, 也是 AIGC 的语料库和资源池。然而, 应当对数据商品化、人与数据深度绑定、数字劳动等保持批判意识, 避免沦为“数据中心主义”。在 AIGC 带来的新闻传播业变化中, 应找寻人的主体价值, 倡导良性的数字交往。

关键词: AIGC; ChatGPT; 数字新闻学; 大数据; 传播政治经济学

中图分类号: G20

文献标识码: A

文章编号: 2096-8418 (2023) 03-0009-08

基于人工智能技术驱动的自然语言处理 (NLP, Natural Language Processing) 应用 ChatGPT, 已经掀起了一场全社会的讨论热潮。作为专门化人工智能领域的里程碑式的产品, ChatGPT 仅在发布五天后就已突破 100 万用户, 并在发布两个月内, 实现月活跃用户突破 1 亿, 成为史上用户数增长最快的消费者应用。迄今为止, 该应用已经迭代至 GPT-5 阶段, 而由 GPT 技术引发的新一轮的内容生产和新闻传播变革正在成为学界业界的关注热点。

不同于高门槛、长周期的专业生产内容 (PGC) 和泛参与性的用户产生内容 (UGC), 学界和业界将这种基于大数据、互联网和人工智能等技术生成的内容称为 AIGC (AI-Generated Content, 人工智能生成内容)。AIGC 是在大模型基础上、基于自然语言处理技术的生成式内容, 其底层逻辑是一个超级规模的大数据结构, 并以此为基础的算法学习和对话式内容生成, 其核心在于由超级规模的数据所驱动, 并通过对抗学习生成模型且不断迭代。

可以说, 在 AIGC 带来的新闻传播活动变革下, 研究和讨论本已十分重要的“大数据”问题具有理论和实践的双重意义。若从更宏观的背景来看, GPT 技术和 AIGC 的发展正是大数据技术已经深度嵌入社会系统之中的表现。从经济社会活动来看, 近年来我国实施大数据战略取得显著成效。从日常生活中的个人移动支付、征信记录查询到疫情防控下的健康码、行程轨迹等数据使用, 大数据塑造着人们的交往行为与观念, 更是成为社会组织、协调、动员的重要资源。然而, 在现实运作中也出现了唯数据、制造数据、“只看数据不看人”等问题, 这些都对社会治理带来了新的挑战。作为一种重要的生产要素和治理手段, 大数据同样是需要被治理的对象。

基于此, 本文所讨论的核心问题是: 在 AIGC 所引发的新闻传播变革背景下, 从新闻传播学的角

度,大数据的角色和功能究竟是什么?大数据对新闻传播活动带来的影响有哪些?我们应当对当前大数据的发展保持怎样的反思和批判姿态,并推动新闻传播活动的良性发展?

一、数字社会环境下关于大数据的基本认识

正如 20 世纪中期以自动化和装配线为代表的福特主义的出现,改变了人们对劳动、工作和社会的理解一样,大数据的诞生也重塑了我们认知世界和理解社会的方式。大数据(Big Data)作为一个概念,最早由美国科学家约翰·马西(John Mashey)于 1998 年提出,以此描述数据量的快速增长带来的四个挑战,即难理解、难获取、难处理和难组织。^[1]随着经济社会和数字技术的发展,尤其是网络社会的来临,大数据在社会中扮演的角色愈发重要,具体可归纳为四种面向。

第一,大数据作为一种中性的技术手段,具有提高经济社会效率(包括传播效率)、提升社会治理能力等功能。自科学主义盛行以来,自然科学的法则被普遍应用在其他领域。而在实证主义研究范式的推动下,数字在人类社会运作过程中的作用越来越大。大数据及其技术的发展,更是提供了前所未有的广度、深度和规模收集和分析数据的能力。在国家和社会治理中,大数据被视为数字建设的核心要素,能够引起“管理革命”,为治理决策提供依据。数据的社会化和社会的数据化相结合,有助于政治、经济、民生、医疗等宏观领域的资源合理配置。

大数据还被普遍运用于突发事件预警和快速响应机制中。新冠肺炎疫情加速了全球社会的数字化生存,数据则以各种形式嵌入了人们的生活。从个体的行程码、健康码到社会层面的感染数据、传播轨迹、核酸检测数据、疫苗接种数据等,大数据连接起个体与社会,疫情传播规律和发展态势得以被精准揭示。从个体的数据化生活来看,伴随着数据管理和应用的成熟,人们通过记事本、备忘录、健康 APP 等各种数字媒介工具,实现对个人数据或家庭数据的自我管理。总的来看,大数据是一种物质性、工具性的存在,社会功能和意义显著,这也是人们对于大数据的乐观认识。

第二,大数据日益成为一种普遍的社会观念,是现代知识体系中一类新型知识(知识资源)。2011 年的世界经济论坛提出,数据作为 21 世纪的一种宝贵资源,是与资本、劳动力相当的新型原材料。此后,大数据概念风靡开来,甚至被认为是信息社会中一种重要的生产力。大数据主要指搜索、汇总和交叉引用大型数据集的能力。^[2]在社会学视野中,社会事实的客观性离不开对数据收集和分析的系统性方法,数据被认为可以客观全面地呈现社会事实,并使得社会科学工作严谨可靠。大数据作为被概念化的社会事实,通过各个学科的转译,在社会运行中越来越不可或缺。近年来,各个学科关于大数据的研究反映出社会思想和研究中的计算转向。^[3]总之,大数据在经济学、管理学、社会学、传播学等各个学科知识体系中的重要性,揭示了大数据本身作为一种知识资源的价值。

第三,从批判的角度反思“数据中心”的社会发展逻辑,重新认知大数据的价值。现代经济社会发展和社会治理离不开大数据的支撑,然而也存在大数据的误用等问题,比如过度追求数据化(流量)、将复杂的社会现实简化为数据等。就新闻传播业而言,大数据可能影响甚至干扰新闻工作者的判断,有的时候过分依赖数据信息,甚至出现了“无数据,别说话”(No data, No talk)^[4]的新闻传播观念。这种“唯数据论”的倾向忽视了制作有深度的、受众所关切的新闻内容所需的时间,也放弃了自身对事实变动准确报道的能力。正如研究者观察到的,大数据并不具有自我解释性,更不是所谓的“客观真理”,而解释数据的方法却可以引起各种意义的争论。^[5]换言之,我们还是有必要反思作为资源的大数据的稀缺性,“大”不一定意味着全面和真实,也可能是冗余和繁杂,因而也有研究者呼吁重思“小数据”的价值。

此外,社会不同利益群体和价值的分化,使得大数据的管理方面也出现诸多问题。诚然,技术本无善恶,但技术作恶却有一个共同特征,即“既有的社会规范无力约束新的技术行为”^[6]。如未能秉承技

术节制原则的“文明码”的开发和应用,就是技术治理的数字化异化和基层治理的简约化动机带来的“滑坡效应”^[7]。

第四,大数据也应当作为被治理的对象,大数据运用中的关键是“数据善用”,即针对数据本身的治理问题。从近年来全球互联网治理实践来看,“治理数据”成为各国的共识。就治理的手段来看,包括个人数据保护、数据隐私等在内的立法和监管手段也在跟进,如欧盟制定的数据安全法规《通用数据保护条例》(GDPR)、美国《加利福尼亚消费者隐私法案》(CCPA)等。我国也出台《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律和行政法规,并持续完善数据监管的标准和规范。

平台(platform)是数据收集、处理与保障数据安全的信息基础设施,也是大数据治理的关键对象。作为世界第二大经济贸易体,我国拥有多个亿级用户的互联网平台。2020年以来,我国开启迄今最大一轮的互联网反垄断治理工作。国家通过立法和监管等综合手段,推进构建大数据治理体系和公平竞争的经济和社会秩序,确保不同主体的数据收集、管理和应用的透明性,以实现“数据善用”“数据向善”的核心目标。

二、数字新闻传播中的基础资源和关键要素

从新闻传播学的视角来看,人类的新闻传播活动离不开数据的支撑,信息的采集、加工、传播和反馈是新闻业运作的基本环节,而数据则构成了信息的基本单元。最早的手抄小报登载的内容除了文字简讯,就是数据,包括商品内容、价格等信息。在数字新闻业时代,不管是算法推荐技术的广泛使用,还是融合新闻生产、智能新闻生产等新闻传播新形态的兴起,大数据越来越扮演着基础性功能,构成新闻传播活动的基础资源和关键要素。

(一) 大数据作为新闻生产活动的资源

首先,在新闻采集环节,数据是信息的基本单位,一定程度上“表征”了客观事实。可以说,新闻最开始的地方就是数据,新闻从业者们通过采集丰富的数据,以保证新闻报道的准确、真实与客观。早期计算机技术引入新闻工作,成为辅助性报道内容(computer-assisted reporting, CAR)。而这种通过使用数据库和人工调查相结合的方式,为后来的“精准新闻”提供了方法基础。但新闻工作者们知道,数据本身并不等同于事实,记者们也需要找出“隐藏”在数据中的故事或者解释,这才构成通常意义上的新闻。当计算机技术和社会科学不断结合,诸如数据新闻(data journalism)、计算新闻(computational journalism)和机器人新闻(robot journalism)等基于大数据的新闻报道形态便不断涌现。

其次,在新闻报道或者“加工”环节,大数据被阐释为文本,用来描述新闻事实及其变动。大数据背后的深层逻辑在于,由量化思维驱动的实证主义取向认识论,使得新闻业偏向数据导向(data-oriented)的报道趋势,而记者也正在成为一种“确定性的信徒”^[8]。正是在这个意义上,数据在新闻报道中的使用越来越频繁。此外,新闻媒体还构建了一套数据指标体系,将数据直接当作反映社会现实的依据。以商业新闻编辑室推出和发布的国家股市指数为例,诸如《华尔街日报》的道琼斯指数、《金融时报》的富时指数和《日本经济新闻》的日经指数等,就被认为是经济社会变动的客观依据,进而在新闻业发展过程中成为一种重要的新闻样态。

最后,在新闻的传播和反馈环节,大数据可以揭示新闻报道是否产生了价值,并将原本零散的、琐碎的新闻报道整合为完整的传播图景。在注意力稀缺的时代,数据的价值就在于可以成为一种简化的指标,便于衡量新闻报道的效果、用户的关注度乃至成为“决定什么是新闻”的价值标准。从传统的电视节目收视率,到社交媒体时代的新闻阅读、点赞、转发和评论量等指标,数据似乎成为传播资源的代名词。当然,数据的过度商业化不免让人警惕和担忧。

（二）大数据作为社会联结的要素

数字新闻中的数据不只是一种抽象的数字性存在，同样具有物质性的内涵。正是在这种物质性的流动中，新闻业得以参与公共世界的建设，并建构了自己的存在。^[9] 在数字新闻生态中，数据作为一种基础要素和物质资源，构建起人与人、人与社会之间广泛的网络关系。

数据联结起各主体之间信息流通。在记者的日常工作中，最重要的任务之一就是采访对象那里获取信息和数据，并建立记者和信源之间的联结。而为了达成社会信任，无论是政府，还是商业机构，都会主动给新闻机构提供数据，保持与公众之间的联结，数据也意味着公开透明。数字时代，记者不仅需要借助数据完成日常性的新闻内容生产，还要基于大数据维护新闻关系、开展舆情管理。此外，为了追求新闻的透明性原则，关于数据的来源和生产过程也会尽量公之于众。在一些情况下，新闻机构还会使用用户生产贡献的数据素材，如新华社的“全民拍”、澎湃新闻的“湃客”等。

大数据本身还是一种交往的方式。在传统的新闻生产中，记者往往依赖于组织机构和精英群体组成的新闻信息网，而大数据则带来了摆脱这种关系束缚的可能。记者可以成为数据的“管理者”和“解释者”。通过数字基础设施和各种传播分析工具，记者得以更好地收集、组织和分析网络中不同的信息流，预测、传播社会事件的过程及其发展趋势。

（三）大数据作为数字新闻业中的行动者

数字时代的新闻生态呈现出一种更为混杂（hybrid）的状态：技术思维和平台逻辑深刻嵌入新闻业的运作之中，人类行动者和非人行动者（算法、文本等）的行为相互交织，带来了新闻传播活动的巨大变化。如果说在传统新闻业时代，数据只是作为一种工具性的应用存在，那么在数字新闻业时代，数据则是一个具有一定的自主性的行动者（actor），在新闻生产环节中发挥十分重要的作用。这不仅是行动者网络理论（ANT）带来的启发，而且是在数字新闻生态下人们对大数据的新认识。

大数据及其系统能够实时、全时地对社会事实的变动进行监测和预警，进而深刻影响新闻生产。作为一种稀缺资源和动态工作安排，时间制约并塑造了新闻生产。在数字环境中，新闻时间除了受新闻工作者和媒体机构影响外，还受制于基于大数据运作的新闻节奏。^[10] 在大数据构建的新闻节奏下，任何新闻机构都不能、也不愿意脱离大数据使用的效率和便捷，个体的新闻工作者更无法在平台、用户和算法之外单方面改变这种数字新闻生产的节奏。

基于互联网和数字平台的传播实践，数据逻辑成功地嵌入数字新闻生产的全流程中。当前，我们讨论的数字新闻学（digital journalism）并不仅仅局限于数据新闻这一狭义的范畴，而是从选题策划、编辑流程到呈现样态、反馈机制、传播网络等方面，新闻工作者都需要符合数字新闻业的规则、惯例、时间节奏等生态原则，要学会“与数字共生”“与数据一起工作”。在这个意义上，大数据作为一个非人的行动者，为新闻工作者带来了新的角色定位、工作内容和交往关系，也塑造了数字新闻业的运作逻辑。

（四）大数据作为 AIGC 的语料库和资源池

ChatGPT 技术的发展和 AIGC 的兴起，使得大数据的重要性愈发凸显。根据 Open AI 2020 年发布的论文，其训练使用了 4990 亿 Token^① 的数据。这些训练语料中，60% 的内容来自过滤后的 Common Crawl^②，22% 来自 WebText2^③，16% 来自各类书籍，3% 来自维基百科。^[11] 而基于大量的数据集和多模

① Token：最小语义单位，单词的信息。即单词内部表示的数据结构形式。

② Common Crawl 是一套数据集。其从 2018 年开始搜集数据，来源包含了博客、网站、维基百科、各国网络社区、政府和大学官网等。

③ WebText：文本数据库，来源于谷歌、电子图书馆、新闻网站、代码网站等。WebText 为 GPT-2 所使用的数据库，ChatGPT 在此基础上做了扩展成为 WebText2 但并未公布其来源。

态模型 CLIP (contrastive language-image pre-training)^[12], AIGC 得以区别于传统 UGC 和 PGC, 实现跨模态融合 (即通过分别提取文本特征和图片特征进行相似度对比, 经由特征相似度计算匹配关系) 的内容生产。

在 AIGC 所带来的影响下, 数字新闻业能够基于更大维度的数据进行新闻内容生产、分发与流通。具体而言, 全自动链条的内容创作与到分发程, 使媒体的生产更迅速和高效, 并能够简化媒体专业人员的复杂任务。^[13] 而基于自动分析用户行为及其接触体验, 媒体能够提供更多个性化的新闻内容^[14]。最后, 通过大数据集的用户行为分析, 帮助内容创造者和市场营销人员能够更好地识别定位用户及其消费模式, 从而改进的最佳方式。^[15]

与此同时, AIGC 也将从多方面对既有的新闻生产模式、分发反馈机制形成全面冲击。在数字时代尤其是 AIGC 变革下, 大数据作为新闻传播活动的基础资源和关键要素, 也将成为各方争夺的场域, 这就需要我们传播政治经济学的视角展开进一步分析和批判。

三、传播政治经济学视角下的大数据及其批判

历史地看, 人类社会在不同的阶段, 随着生产力的发展带来了新的资源配置形式。在农业社会, 土地和劳动力是核心的生产要素。在工业社会, 资本、石油钢铁等工业原材料、机器则是不可替代的生产资源。当人类社会进入数字时代, 在以 ChatGPT 技术为代表的新一代人工智能发展背景下, 大数据日益成为社会经济活动的重要资源, 影响着社会生产、分配、交换和消费的各个环节。

(一) 资本的垄断: 被“商品化”的大数据

就互联网企业而言, 数据资源不仅是实现价值创造的核心, 更是“权力跃升”的关键。但是, 数据的收集、管理、存储、分析和应用的过程却一直是一个“黑箱”。2019 年, 学者肖莎娜·祖波夫 (Shoshana Zuboff) 通过观察美国互联网商业巨头经济模式, 创造性地提出监视资本主义 (Surveillance Capitalism) 的概念。^[16] 有学者批判道, 在数据经济的运作中, 用户的个人数据及隐私被资本所“掌握”, 活生生的人被用作公司企业的敛财工具与生产资料。^[17] 具体来说, 通过将个体行为数据信息作为免费“原料” (raw material), 互联网平台公司或出售给第三方公司直接盈利, 或通过算法进行产品预测和生产, 最终转化为市场中的“期货商品”。由此完成将数据的“使用价值转化为交换价值”这一商品化过程。

传播政治经济学一般认为, 使用价值在资本主义体系中变得货币化, 这使得商品能够在市场上出售。^[18] 虽然单一数据的使用价值有限, 但大数据却让足量异质化的数据集产生较高的交换价值。互联网平台通过挖掘数据之间的关系, 将再次提升数据的商业能力。为了实现资本的快速积累与增值, 数据的自由流动则会受到一定的限制。平台之间会设置壁垒防止用户的流动, 比如用户面临二选一服务, 不同平台之间支付跳转和外链打开受限等。平台通过实施保护主义, 实现资源的封闭和占有, 控制了数据资源的商品化能力。

总而言之, 在数据的商品化过程中, 平台或互联网企业通过收集、分析和买卖个人数据来牟取超额的利润, 将受众商品和数据商品合二为一, 在一定程度上垄断了大数据资源。这就使得我们一定要注意 AIGC 背后的超级规模数据集的商品垄断问题。

(二) 数据的非扁平化: 人与大数据的深度绑定

作为一种传播资源, 数据看似是无差别的、扁平的, 但其背后却是现实的社会阶层和社会关系。在平台与用户之间, 这种关系是非扁平化的, 平台拥有为用户进行精准画像的能力, 并进一步通过提供差异化服务固化了这种社会关系, 在将人们与数据深度绑定的同时, 实现层级化的数据关系。

平台的发展过程经历了从技术服务者到数据控制者的角色转换。在法律规定范围内, 为了扩大自

己的功能和效益,平台以“地推”和“让利”的形式吸引大量的用户群体注册和使用。每个用户将拥有一个认证账户,个体信息被纳入平台数据库中。用户之间的互动行为和数据将被永久储存。而当用户加入平台之后,需要接受一定条件的服务条款限制,而平台这一脱离物质实体的空间产物则成了一种自治属性的“领地”。诚然,用户在平台之间具有跳转和流动的机会,但过往储存的数据却很难流动。平台也会使用各种手段来增强用户黏性,比如通过价格战、补贴战等手段拉拢用户和商家,通过满单奖励拉拢外卖骑手等劳动者等,由此形成一种临时性的结盟。这种利益共谋的身份,深度绑定了每一个平台使用者。

在此意义上,大数据的兴起与互联网空间中的层级化或者圈层化过程密切相关。基于大数据,平台通过算法等技术手段,将用户身份进行标签化、层级化处理,由此大数据再次呈现出现实社会中的社会分层图景,具有一定的“象征权力”。同时,现实中的社会结构和社会关系也决定了个体对数据的控制、占有和分配状况,这反映了不同社会阶层在数据的使用和自我保护等方面是有所差异的。

(三) 劳动“异化”:从受众商品到数字劳工

在大数据的生成过程中,个人数据变成了一种抽象的劳动商品。不同于传统的生产劳动,互联网的数据劳动大多是用户自发的行为,互联网用户不是被动的受众,而是数据生产的主体。在加拿大传播政治经济学学者达拉斯·斯迈思(Dallas Smythe)看来,媒体受众在消费媒体内容时观看广告,实际上是在他们所谓的免费休闲时间,通过观看媒体——尤其是广告——为广告商服务。^[19]简言之,受众收看广告行为是一种商品化劳动。当然,受众商品论并非仅指用户收看广告的行为,而是受众的注意力这一稀缺资源,其本身就是一种排他性的商品。进入互联网时代,平台、广告商与用户之间的关系发生了新的变化,用户数据成为三者关系的重要纽带。

早在2011年,针对谷歌企业受众商品的一项研究证明,没有用户的集体劳动,搜索引擎中的广告剩余价值就不复存在,用户的每一次点击和搜索行为都会被定位,进而提升其交换价值和广告投放。^[20]如果说传统意义上“受众商品”的交换价值取决于“人口统计学数据”,那么数字时代则指的是点击、评论等交互行为所产生的数据。用户在互联网使用过程中,无论是工作时间,还是休闲时间,无论是通过平台开展工作,还是社交活动,都在源源不断地为平台生产数据,而这些数据则进一步成为具有交换价值的劳动商品。在大多数时刻,用户沉迷于互联网中,并未意识到这是一种数字劳动,无形中为平台创造了巨大的经济价值。

由于数字技术的快速发展,用户现在可以享受到多种多样的信息和便捷的服务,然而数字劳动的形式以及相关的量化技术和计数指标却更加复杂,日益打造了一套数据的“牢笼”。数据成为平台激励政策的变现手段,还通过隐形货币化(如打榜、投票等)测量用户的数字劳动,并在平台内流通和兑换。在这一过程中,互联网平台通过各种形态的数字劳动,完成了数据资本的积累和再生产,“为了数据而数据”就是另一种典型的异化劳动现象。

可以看到,大数据的产生和应用全方位地渗入用户日常行为之中,从隐私信息、在线操作到内容生产,“数字劳工”持续不断地为数字资本的累积与增值提供数据,由此资本和平台的控制权力得以巩固。因此,我们应当对数据商品化、非扁平化和数字劳动等问题保持警惕和批判意识,通过加强法律法规、健全平台自律、提高个人数据素养等方面,以避免大数据所带来的极端化状况。近年来,“治理数据”“数据向善”等理念受到了更多的关注,“人—数据—平台—社会”之间的关系亟需重构,我们需要迈向“数字交往”的新维度。

(四) 摒弃“数据至上”:AIGC发展中的主体性问题

在AI技术的演进中,AIGC内容共经历了三个阶段:基于实体孪生阶段、基于学习创作阶段和基于实时自主生成阶段^[21]。但当Chat GPT介入知识生产时,其模糊了知识生产中“人类要素”与“非人类

要素”的界限, 形成了大量“杂合知识”, 人类的生存境遇也将随着知识生产的此种变化而改变。^[22]

在 AI 技术的作用下, 一种全新的数字人正在产生, 而“虚拟化、智能化、外脑化”便是其发展方向。这种意识复杂化的人, 以及被数字技术“交往化”的人, 既是技术的演化, 也是新人的初生。^[23]但正如前文提到的, 大数据的价值在于提供了感知、预测和治理社会世界的新思路和新模式, 其核心就在于通过数据驱动的决策替代经验性决策。然而, 当越来越多决策角色被技术和算法所取代, 个体难免呈现出客体化的趋势, 即人的主体性的丧失, 其为生活赋予意义的能动性被剥夺, 可能沦为被平台和信息支配的个体。^[24]这就是数据至上的观念和实践所带来的极端状况。

对大数据的追捧也让人们忽视了小数据的价值, 有时还违背了数据选择、收集和使用的道德伦理原则。实际上, 在新闻传播活动中, 以数据为驱动的智能媒体实践已迭代多年, 从最早的机器人写作、智能剪辑、算法推荐到近年来爆火的 AI 主播, 数据聚合和信息可视化作为新闻生产的一种重要方式。在此过程中, 大数据常常被视为中立、客观的工具, 然而却忽视了将新闻作为一种交往的追求。以 Chat GPT 为例, 在“信息论(概率)”和“控制论(反馈)”的影响下, 其所生产的信息内容是缺乏反思性的逻辑关系对既有知识的概率联结和组合^[25], 而这种拼凑式内容创作也对现有“原创作者”观念和制度带来了挑战。当新闻的选择、决策和判断交由数据、平台和算法, 而非凭借记者的主观能动性时, 人在新闻传播中的主体性地位将遭受冲击。数据并不能代替新闻传播主体的理性行为, 更无法揭示新闻传播活动中复杂的情绪、社会心态和社会互动。

四、结语: 从基础资源到数字交往

AIGC 为人类社会的文化与艺术信息的表达、存储和传播提供了一种新的形态, 也将引发教育、传媒、影视、电商等社会诸多领域的变革, 影响数字时代人们生活的方方面面。本文认为, 在以 Chat GPT 为代表的 AIGC 技术变革中, 大数据是数字新闻传播活动中的基础资源和关键要素, 一定意义上还是能够自我再生产的“行动者”和“新新数据”。当然, 就其本质而言, 大数据并不是自然产生的, 而是一种历史建构的产物, 并影响着人类社会发展的现实走向和逻辑路径。

在大数据广泛运用尤其是 AIGC 日益发展的语境下, 数字新闻业愈发离不开大数据的采集、整合与应用, 但更应明确的是: 新闻传播主体之间的关系核心仍然是人的交往关系, 新闻业不能陷入数据中心主义的泥沼, 而是要坚持以人为本的传播理念与实践形态, 即从单纯的信息传播到更高维度的沟通对话以及关系的维系。^[26]

质言之, 在数字生活世界, 个体既是一种生命实体, 也是一种符号化和抽象化的存在, 不同主体间通过数字传播技术, 建立交往网络与社会互动。在被数字技术全面中介化的当下, 数据成为人与人之间建立关系的重要连接点, 但这种连接最终是要靠人来实现的, 而非数据本身。人们开发和使用大数据及其技术, 根本目的是建立复杂、有序且有意义的人际联系。“数据拜物教”只是一种物质交往形态, 而基于大数据展开更深层次的“数字交往”, 突出精神交往的维度, 则是一个更高的境界。对于新闻传播活动而言, 数字交往意味着在大数据应用中找回人的主体性, 将大数据作为连接人与世界的中介角色。

参考文献:

- [1] 谭建荣. 大数据思维与新基建发展 [EB/OL]. 中国信息化周报, 2021-10-11.
- [2] Boyd, D. & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15 (5): 662-679.
- [3] Burkholder, L. (2019). *Philosophy and the computer*. New York: Routledge.

- [4] Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1 (1): 1-12.
- [5] Bollier, D. & Firestone, C. M. (2010). *The promise and peril of big data*. Washington, DC: Aspen Institute, Communications and Society Program.
- [6] 邱泽奇. 技术化社会治理的异步困境 [J]. 社会发展研究, 2018 (4): 2-26+242.
- [7] 郭春镇. 对“数据治理”的治理——从“文明码”治理现象谈起 [J]. 法律科学 (西北政法大学学报), 2021 (1): 58-70.
- [8] Anderson, C. W. (2019). Between certainty and uncertainty: The historical, political, and normative contexts of inferential journalism claims. *Journalism & Communication Monographs*, 21 (4): 358-361.
- [9] Anderson, C. W. & De Maeyer, J. (2015). Objects of journalism and the news. *Journalism*, 16 (1): 3-9.
- [10] Ananny, M. & Finn, M. (2020). Anticipatory news infrastructures: Seeing journalism's expectations of future publics in its socio-technical systems. *New Media & Society*, 22 (9): 1600-1618.
- [11] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P. & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33: 1877-1901.
- [12] Radford, A., Kim, J. W., Hallacy, C. et al. (2021). *Learning transferable visual models from natural language supervision*. Paper presented at the Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research. Retrieved May 2, 2023, from <https://proceedings.mlr.press/v139/radford21a.html>
- [13] Thurzo, A., Strunga, M., Urban, R., Surovková, J. & Afrashtehfar, K. I. (2023). Impact of artificial intelligence on dental education: A review and guide for curriculum update. *Education Sciences*, 13 (2): 150-152.
- [14] Rathore, B. (2023). Integration of artificial intelligence & its practices in apparel industry. *International Journal of New Media Studies (IJNMS)*, 10 (1): 25-37.
- [15] Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, 10776958221149577.
- [16] Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs.
- [17] 符豪. 21 世纪美国左翼学者对监视资本主义的批判研究 [J]. 马克思主义与现实, 2020 (5): 135-140.
- [18] [加] 文森特·莫斯科. 传播政治经济学 [M]. 胡春阳, 等译. 上海: 上海译文出版社, 2013: 168-170.
- [19] Smythe, D. W. (1981) On the audience commodity and its work. In Durham, M. & Kellner, D. M. (eds.). *Media and cultural studies: Keywords*. Hong Kong: Blackwell Publishing, 230-256.
- [20] Kang, H. & McAllister, M. P. (2011). Selling you and your clicks: Examining the audience commodification of Google. *Triple C: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 9 (2): 141-153.
- [21] 李白杨, 白云, 詹希旎, 等. 人工智能生成内容 (AIGC) 的技术特征与形态演进 [J]. 图书情报知识, 2023 (1): 66-74.
- [22] 姜华. 从辛弃疾到 GPT: 人工智能对人类知识生产格局的重塑及其效应 [J]. 南京社会科学, 2023 (4): 135-145.
- [23] 杜骏飞. 何以为人? ——AI 兴起与数字化人类 [J]. 南京社会科学, 2023 (3): 76-85.
- [24] 孙国东. “算法理性”的政治哲学检视 [J]. 厦门大学学报 (哲学社会科学版), 2021 (2): 23-33.
- [25] 邓建国. 概率与反馈: ChatGPT 的智能原理与人机内容共创 [J]. 南京社会科学, 2023 (3): 86-94+142.
- [26] 涂凌波, 虞鑫. “新闻价值”学术对话: 数字新闻语境下的变革及其未来 [J]. 青年记者, 2022 (9): 12-17.

[责任编辑: 高辛凡]