制衡还是共生：生成式人工智能治理的技术校准与行动框架

杨　雅　苏　芳　章雪晴*

【摘　要】生成式人工智能的发展突出了技术的加速度与社会治理框架的不协调问题。本文以生成式人工智能对社会结构的突变性影响为切入点，分析以数据安全为代表的技术风险、技术利维坦和隐私安全等社会风险以及以认知思维构陷为代表的决策风险等突出问题，在技术权力置换与社会格局重组中寻找生成式人工智能技术校准的基准点，并以中观与微观数智素养培育为行动起点，以期在“有形边界”之外，为生成式人工智能发展提供“无形边界”与伦理素养规范的治理逻辑与行动框架。具体来说，首先，在宏观层面，需要构建一个全球性的治理框架，以应对生成式人工智能的全球性挑战，在跨国际的治理合作中构建一个安全、公平、可持续的全球性生成式人工智能治理体系，实现广域的数字化联合。其次，在中观层面，需要建立健全制度体系，以规范生成式人工智能的研发、应用和推广的政策法律等“有形边界”，同时进行社会层面的道德规制，找准技术校准点，为保障该技术的健康发展提供更有益的社会环境。最后，在微观层面，需要关注个人的惯习和认知，以无形之力搭建起人类与生成式人工智能技术之间的认知边界、信任边界和文化边界，以期在个体层面上提升人们对生成式人工智能的认识和理解。

【关键词】生成式人工智能；无形边界；技术校准；社会治理；数智素养

DOI:10.16775/j.cnki.10-1285/d.2024.02.002

随着人工智能、区块链、深度学习、物联网等新一代数智技术的集成迭代与扩散，在深度媒介化和网络化的社会进程中，生成式人工智能技术不断融合创新，影响社会关系变革，重塑人类劳动方式。个体得以从重复劳动中抽身，投身创造性的劳动。数智技术与人的“共生”带来人与社会发展的价值优势，不过，生成式人工智能的发展也可能带来亟须“制衡”的伦理挑战。诸如数字不平等和信息层面的失序和伦理风险，导致虚假宣传、误导欺诈、意见操纵以及仇恨引导等信息的传播①；人类对大语言模型等技术过度依赖的可能性，降低批判性思维与决策能力②，形成智能时代

* 杨雅，北京师范大学新闻传播学院副教授，博士生导师；苏芳，北京师范大学新闻传播学院博士研究生；章雪晴，北京师范大学新闻传播学院硕士研究生。

① Weidinger L, Mellor J, Rauh M, et al. Ethical and social risks of harm from language models, *ArXiv.Org*, 2021.

② Dergaa I, Chamari K, Zmijewski P, et al. From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 2023, 40(2).

的沉默螺旋，造成公共领域各主体话语权的失衡。

因此，面对技术的加速度，亟须思考与判别生成式人工智能技术的效能与风险，实现社会治理和人工智能的负责任创新，即“创新＋责任”，不仅仅是促使创新的升级，而是从根本思路上重新看待技术创新的伦理和社会层面影响，使得“技术和科学进步正确融入社会”①。在技术的“摇摆”过程中锚定技术校准的基点，在技术的权力“置换”过程中寻找治理逻辑的突破口，在法律政策的有形边界之下，探索个体认知和数智素养的“无形边界”，构筑生成式人工智能的未来治理逻辑与行动框架。

一、摇摆的钟：生成式AI的技术效能与社会风险

数智媒体时代，技术的“钟摆”始终在正向的应用效能与负面的社会风险之间摇摆，新技术动力的社会效能与逆效应并存，技术的迭代也带来了“涌现性的问题域”，使得技术与社会之间的互动关系更加复杂微妙，人与技术的关系也经历着从驯化到协作与共生的交互过程，且对现有社会的文化肌理、社会逻辑和认知框架产生重要影响。

（一）技术层面：技术稳健、数据可信与“机器偏见”

目前大语言模型（以下简称大模型）在起步阶段，可能存在技术发展中的稳健性不足、模型可解释性低、算法偏见等问题。稳健性不足，可能导致训练数据的偏差、模型的错误推理等，目前大模型的鲁棒性主要体现在提示词和对话任务等层面；可解释性是指模型的透明性，当人类难以理解模型的决策过程时就可能出现信任危机和安全问题；算法偏见，则有可能出现在数据来源不均和语料库缺乏代表性的情况下，对于特定的人群、种族、地区、职业等产生机器偏见。

因此，“公平性，隐私性，可解释性”，是当前在人工智能伦理研究领域的核心议题②。在通向“可信且可解释的人工智能”过程中，对齐（alignment）成为人工智能系统与人类意图和价值观匹配的目标，确保智能生成内容符合稳健性、可解释性、可控性和道德性等要求，对人类和社会发展有益。

（二）社会层面：数字鸿沟、机器遗忘与“技术利维坦”

在社会风险方面，需要防范技术所导致的数字鸿沟和隐私让渡问题。数字鸿沟是指社会公众适应大模型等技术时产生的接入沟、使用沟等信息不平等现象。在生成式人工智能时代，由于提示策略和用户大模型素养的差异，数字鸿沟也出现新的表现形式。同时，在大模型的技术知识掌握过程中也可能存在“主客观知识沟”的现象③，即人们在大模型相关研究领域的主观知识和客观知识之间，也存在实际感知上的差异，倾向于错误校准或者是对所掌握的知识“过度自信”；或者正相反，由于这种偏差性的认知，可能会在人内传播中产生无形的知识沟，以及主客观知识不匹配、认知失调带来的所谓“面对技术的不自信”。

同时，技术通过掌握人的隐私数据的让渡，而积累起权力的“原始资本”。算法作为一种新型媒介，也可能导致社会权力关系的重构。随着技术的自主性发展和涌现，最终将摆脱人类控制生成

① Blok V, Lemmens P. The emerging concept of responsible innovation. three reasons why it is questionable and calls for a radical transformation of the concept of innovation. // B. J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications.* Springer International Publishing Switzerland, 2015: 19-35.

② 郁建兴、刘宇轩、吴超：《人工智能大模型的变革与治理》，《中国行政管理》2023第4期。

③ 韦路、秦璇：《主客观知识沟：知识沟研究的新方向》，《新闻与传播研究》2023年第2期。

新的权力，而社会化技术权力的出现可能导致“技术利维坦”的出现，凸显“人性”与“物性”的矛盾[①]。既往字母表的出现驯化了人们的思维方式，而大模型语言的出现也可能导致新的思维习惯和语言特征，通过大量模型数据训练不断逼近人类的意识和创造力，对人类的主体性造成挑战，并由此建构起的新的技术权力惯性。不过，人与技术最大的差异在于自反性、意向性和创造性，在技术与社会发展中凸显人的主体性、真正借由技术的发展促进人的升维，超越时空的界限而发展人的自由度，也成为当下的重要议题。与此同时，近期“机器遗忘”（machine unlearning）的概念升温，这一概念旨在寻找一种技术干预的方式，让大模型在保持较高预测能力前提下，“遗忘”掉训练集的敏感数据，从而使得必要的知识产权与个人隐私权得到保护[②]。

（三）决策层面：信息冗余与认知思维构陷

技术与社会的互动过程集中体现在人类的认知和决策方式上，生成式人工智能的发展也可能使得人类面临认知和决策的风险。在决策风险中，“人类世”中技术的突变带来了信息的熵增，而人类的信息搜寻方式也从自主搜寻转变为与机器的人机对话和人机共创。机器实现了信息的搜集、把关过滤、重新组织，再到个性化分发、匹配受众的各阶段，即在信息整体传播“传递”的过程中，扮演了“把关人”的角色。同时，以对话交流等人际沟通的方式，创造一种模拟共情、模拟多轮对话聊天的传播“仪式”，在潜移默化中影响人的认知思维和行为决策。

生成式人工智能可以在多轮对话中自主“涌现”生成新的内容，这可能对人的认知产生连续的不确定性影响。例如，在人机对话过程中获取使用者的个性特征、态度立场、价值取向、文化属性等特点，并加以调试和顺应生成个性化内容，而这些内容有可能对个人形成信息茧房效应，减少其对信息的核查的警觉意识。在认知方面，生成式人工智能也可能通过认知资源有限性原理，对个体进行认知习惯的塑造、认知信息渗透和价值观的重塑，也即“认知思维构陷”，例如通过情感传播和模因传播的方式来吸引注意和流量以及进行认知渗透。

二、技术校准的基准点：技术权力置换与社会格局重组

在讨论生成式人工智能的伦理和边界规则之前，有必要找到其伦理和边界规则校准的基准点。大模型的素养和伦理规范背后具有结构性的动因，而所谓“科技向善”中的“善”也是一种关系范畴。因此，在伦理校准的底层结构中需要包含结构性和关系性的深层逻辑作为支撑。如前所述，生成式人工智能作为一种革命性的技术，也是一种新兴的生产力，将带来整个社会全方位的权力格局重组，具体表现在资源配置、信息传播范式与社会秩序、社会治理等层面。

（一）权力格局重构：知识与数据、算力、算法成为关键生产要素

权力结构形成了社会阶层和社会结构，在“国家—市场—社会”中形成三元离散性力量的平衡。其中，国家是一种组织形式，市场代表商业化力量，社会则包括公共领域和私人领域[③]。从时间线程来看，在智能化、网络化、赛博格化的过程中，生成式人工智能等新技术的发展，将与石器、青铜器、蒸汽机和计算机等共同成为人类社会变革的生产工具，在新一轮康波周期中渗透到各

① 袁超：《“技术利维坦”的傲慢与偏见——以“社会化技术权力”为中心的讨论》，《人文杂志》2021年第8期。

② Xu H, Zhu T, Zhang L, Zhou W, Yu P S. Machine Unlearning: A Survey. *ACM Computing Surveys*, 2024, 56(1).

③ 刘金河：《权力流散：平台崛起与社会权力结构变迁》，《探索与争鸣》2022年第2期。

结构力量之中形成关系重组，打破原有的均衡博弈局面。

从空间角度来看，周期的运转中伴随着社会格局和秩序的演化。从权力结构的角度来看，技术重塑了知识结构、生产结构和金融结构。由于ChatGPT等大语言模型本身是一种信息重组和辅助人类进行知识生产的工具，因此权力重组过程中，知识效应的权力化也相应地扩展了资本的权力。一方面，知识作为生产要素，成为生产财富的必要手段；另一方面，知识的建构体现出资本的支配特征，知识垄断将导致社会财富和资源倾向于商业性科技巨头，增加其政治性权力。这种“技术置换权力”的机制，使得其既获得了技术进步带来的收益，同时也获得了社会话语和正当性的双重权力[①]。不仅如此，技术公司的内生权力与外部溢出效应、产业格局的聚合与分散，可能使得社会按照技术和数字化的格局重组。生成式人工智能技术和数据要素的依赖抑或自主，也影响着未来国家社会治理甚至区域和国际关系。

基于数智技术在时空维度的重组效应，在传统意义上的接入沟、使用沟与知识沟之上，在国家、社会、个人等各层面，可能出现新一轮的差异与鸿沟，即“AI鸿沟”或“智能鸿沟”。在国家层面，技术资源的富有国与资源贫乏国之间将出现鸿沟[②]；在个人层面，将出现人与机器之间的认知鸿沟，以及由于使用能力和素养等差异所导致的认知鸿沟。例如文生视频软件“Sora”作为场景媒介，将虚拟观看维度提升至在场实践维度，则会在个体层面延展使用者的认知边界。

（二）秩序格局转换：人工智能推动社会信息秩序升维

数智技术的变革推动信息传播秩序的更迭，以及社会连接和重组方式的革新。知识与数据成为关键生产要素，原有社会秩序和组织方式无法适应新的信息传播技术范式，也将在社会关系和社会分工方面进行重组。在此基础上，以技术变革为重要动因，在信息秩序的中介下，社会治理的模式也亟待转变。

在智能信息传播范式方面，主体由社交媒体转向智能体，信息传播方式也从大众传播走向数字智能传播。有学者将Web1.0时代，以大众传播媒介为主体、以现实空间为基础、以闭环价值链和有限内容为特征、信息可读但无法编辑与拥有的有序化生产，概括为“大教堂式”的传播模式；同时将Web2.0时代平台型媒体为主体、以用户驱动为基础、去中心化的即时信息传播，概括为“大集市”传播模式[③]。在可预期的Web 3.0时代，基于视频自动化智能生成的媒介，去中心化程度逐步加深，而信息的熵增、冗余与信息失序的缺陷将在区块链逻辑的嵌入后逐渐减少，智能传播将沿着去中心化、面向用户开源、鼓励用户掌握有价值信息且通过价值链创造的“生态圈”传播模式发展。

（三）治理模式迭代：“生态圈”传播模式与去中心化治理的探索

生成式人工智能时代，“生态圈”式的传播模式将社会秩序进一步打破重组，呈现出沉浸共在、去中心化和再网络化的特点。人的自由度进一步提升，也即人们摆脱重复劳动，更进一步实现“自由人的联合体”[④]，在这种社会形态中，主要的组织方式是去中心化的自治组织（decentralized autonomous organizations，DAO）。数据公地（data commons）的概念提出，数据作为一种公共

① 梅立润：《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）》2023年第1期。

② 郭全中、张金熠：《作为视频世界模拟器的Sora：通向AGI的重要里程碑》，《新闻爱好者》2024年第4期。

③ 方兴东、钟祥铭、张权：《“守门人”的守门人：网络空间全球治理范式转变》，《湖南师范大学社会科学学报》2023年第1期。

④ 《马克思恩格斯全集》（第30卷），人民出版社，1995年版，第107-108页。

物品和集体资源，用户可以自主贡献、获得和使用[①]。由此，在算法和机器伦理下逐渐形成信息化秩序，并受到计算逻辑、数据、算法和算力的影响，形成数字化联合[②]。与此同时，社会分工模式也可能进一步发展。人工智能技术的渗透，改变的不只是个人的工作逻辑，而是产业总体的协同和组织模式，越来越多的组织协同会进入整体自动化的体系，机器与机器之间形成庞大且复杂的协作网络；同时这还体现在人工智能对于行业顶尖人才的赋能，供需和分配系统在一定程度上会被重塑，跨行业的溢出效应也可能带来普惠价值[③]。

落脚于治理模式层面，由于信息传播与社会秩序变化的主要逻辑是从中心化到去中心化的变革，与之相对的，治理模式也将从自上而下中心主导模式，发展到纳入自下而上的用户主导模式，形成多源流、多通道、多利益相关方的合力。伴随着社交媒体和智能媒介出现，当前的治理模式也同样逐渐从中心化治理模式向去中心化的治理模式过渡。不过，当知识、数据、算力和算法成为新的生产资料后，有学者提出，新质生产力也号召“看得见的手”对技术自治模式的把关[④]。理想的社会秩序和治理模式既非是大教堂般的一维管理，也非大集市般的冗杂无序，而是在此基础上的可协商、可理性交往的公共领域的再构建。这种理想治理模式，需要政府、社会、个体的投入，减少技术资源不平等带来的数字排斥问题，在技术校准中加入伦理和规范的要素，增强国家之间的协同合作效力等。

综上，当前社会阶段已经处于新的康波周期中。生成式人工智能技术发展将带来现有社群秩序的解构，以及社会权力格局和治理方式的重构。因此，对于技术的伦理边界规则的探讨，也需要以上述的宏观视野为基础，在技术的突变作用中，观照中观制度和微观个人惯习，商讨大模型使用的“无形边界”。

三、大语言模型使用的“无形边界”和伦理素养

生成式人工智能技术越来越呈现出“类人化”特点，但这并不表示技术是万能的，其也可能带来某些错误预判、加重算法偏见，甚至引发社会风险的问题。因此，在支持鼓励人工智能产业发展的过程中，国家政策决定者、产业管理和技术人员、个人用户等多利益相关责任主体需要通力合作，在探索中逐渐清晰勾勒出大语言模型发展的动态适度的边界。

媒介技术的发展深刻地受到社会环境的规制与影响。当前我国针对生成式人工智能的政策环境，呈现出包容审慎、分级分类监管的态度与基调，国家连续出台了一系列行业政策与法律法规，引导AIGC产业向可靠可控、规范应用发展。在数据服务和数字基建方面，提出加快形成“新质生产力”，以新产业、新业态和新模式快速涌现为重要特征，构建起新型社会生产关系和社会制度体系[⑤]，提升我国的竞争优势。

一方面，“有形边界”的重要性显而易见，需要运用更加综合性、通用性的政策法规力量保障个人合法权利、促进产业健康发展、维持社会稳定的规则边界；另一方面，同样需要关注生成式人

① McDonnell J (ed.). A New Deal on Data, *Economics for the Many*, London: Verso, pp.164-173.

② 张兰、李红权：《人工智能：由现代社会向马克思“自由人联合体”过渡的技术路径》，《学术探索》2022年第3期。

③ 白果：《失业、分配不平衡和结构性转变：人还能否“卷”过AI》，经济观察网，http://www.eeo.com.cn/2023/0722/598986.shtml，访问日期：2023年12月22日。

④ 方兴东、钟祥铭、张权：《“守门人”的守门人：网络空间全球治理范式转变》，《湖南师范大学社会科学学报》2023年第1期。

⑤ 于凤霞：《加快形成新质生产力 构筑国家竞争新优势》，《新经济导刊》2023年第10期。

工智能技术在伦理规范、主体使用中的“无形边界”，在个人、社会和文化层面，搭建人工智能技术的认知边界，以及人机互动的信任边界与人类社会的技术文化边界。

（一）“无形边界”：认知、伦理与素养的搭建与培育

在大模型治理中也需要考虑到看不见的无形边界，例如惯习、认知和素养等既有文化与思维习惯形成的边界，从关系、情感和信任等非理性要素的层面，影响个人对大模型的认知以及人机互动行为的塑造。

1. 个人认知边界：用户思维、认知系统与媒介素养

无论是分析式人工智能还是目前正在蓬勃发展的生成式人工智能，智能技术逐渐开始反过来影响和改变着人类的思维方式。在问题解决和决策方面，人工智能技术通过强大的运算能力和数据分析能力，帮助人们快速获取和处理大量信息，提高决策效率，优化解决问题的方式；在创造力与创新方面，技术的便捷性能够节省繁杂重复的劳动，辅助人类集中进行思维的创作，激发人类的创造力和创新精神，在沉思与知觉猜想的空间里，形成“想象力”的探索，设计新产品或发现新颖的科学理论。在我们对于生成式人工智能技术带来的工具性、创意性具备进一步深入感知的当下，更应思考大模型应用于日常生产、创意劳动中的角色定位与个人主动使用的边界。

在人工智能所生成的混元视域中，虚拟与现实之间的边界逐渐模糊，形成虚实相生、虚实相融的生存环境。在人际传播以及人机互动中，生成式人工智能也日益融入人类的精神交往与社会交往。例如，运用生成式人工智能聊天机器人频繁进行人际咨询以及情感代偿，一定程度上可以作为日常情感实践的“前测”和“训练”。不过，过度依赖工具进行回答交流，也可能致使人际交往能力降低，导致人们忽视真实人际关系中切实可感的交流，甚至降低沟通实践和社交的能力。

此外，场域影响着人们的观念行为，随着媒介丰富度和在场感的提升，人们的信息获取逐渐从经验获取、媒介获取转向技术获取。技术辅助人类进行认知，个体也面临“勤思”还是“疲怠”两种认知模式。生成式人工智能可以更加智能化地回答我们的问题，完成我们制定的任务指令，不过结果的简单易得有可能让人们忽视独立思考判断以及深入研究和溯因的过程。同时，多模态信息类型的丰富改变着传统信息生态，信息过载无形中也可能增加个体的认知负荷。

2. 社会文化边界：大语言模型使用的信任、意识与向善赋权

回顾人工智能的技术发展，其社会角色经历人类的个人助手、专业顾问、创新伙伴甚至全球公民的重要转变，生成式人工智能技术未来发展的巨大潜力也为元宇宙等数字空间的未来图景预示了无限可能。在生成式人工智能不断类人化、普及化、通用化的进程中，人类与生成式人工智能技术之间的关系如何走向共生和互构，大模型使用中人类对于自动生成信息内容的信任边界等问题，都反映了对于人与人工智能技术关系本质的思考。

面对生成式人工智能所表现出的类似“有意识”的行为，人们也开始思考它是否真的具有内在的主观经验。曾经精神哲学领域关于“哲学僵尸”的思想实验，探讨了人类如何能够区分真正具有意识的生物，以及仅仅是在行为上模仿意识的生物。当前的机器程序识别方法中，反向图灵测试是其中之一，防止计算机程序冒充人类，保护网站和在线系统的安全，例如安全验证机制验证码（Captcha），通过要求用户完成某个只有人类才能完成的任务，从而证明他们不是机器人。作为一种评估计算机智能的方法，其可以在实践中直接地预防生成式人工智能可能带来的负面影响，也激发了诸如功能主义、全球工作空间等理论层面的发展。同时，多重现实模型、意向性和元意向等观点也为人工智能的识别提供了新视角，我们不仅需要深入理解人工智能和意识的本质，还需要重新

审视人工智能的设计原则、科技伦理与向善赋权。

（二）大语言模型伦理标准：技术的社会规制尝试与负责任创新

生成式人工智能技术作为非人因素（non-human）的行动者主体，也参与到“五际社会伦理关系维度”，如国际、群际、人际、代际、超人际的社会网络建构与关系权力争夺之中①。过去的一年间，生成式人工智能的涌现式发展为全球科技界甚至各行业提振了信心，也有一些担忧的声音如“ChatGPT取代人类、LLMs助推欺诈、打开AGI潘多拉魔盒”等。大模型开发与应用方也因此推出开发AI反馈强化学习（RLAIF）、减少危害内容生成、应对模型偏见等措施来实施治理②。除去技术层面，围绕生成式人工智能伦理风险的讨论还集中在内容层面，如极端与仇恨语言、传播误导性或者虚假信息、侵犯个人隐私和知识产权、恶意使用及滥用技术、资源不均与数字不平等诸多问题。从道德伦理学的角度，这些风险在不同程度上也违反了现有道德体系中的既有准则。例如，偏见和资源不均违反正义准则；误导性或虚假信息违反美德伦理学中的正当准则；仇恨语言违反关怀伦理学中的理念；而知识产权和隐私侵害则违反了效用主义的理念③。遵循负责任发展的准绳，相关治理主体必须着眼于大模型智能带来的风险与伦理问题，努力做到将大模型等强智能体与人类的内在道德价值观相对齐，研究者和开发者也应该采取积极的行动来确保大模型的负面影响最小化，确保大模型的发展能够最大程度造福全人类。

在伦理的标准构成中，也需要构建“技术—生物—社会行为体”的治理框架，以应对“技术—文化”场域（techno-cultural fields）多元行动者混合的空间，即针对生物行为体、机械行为体与社会行为体界定其责任，以及其伦理上的“生的权利”与“死亡（废弃）的权利”④。在此层面上，“负责任创新”这一概念的提出，可以出于技术伦理层面的可接受度、发展的可持续度、社会满意度的全面考量，充分发挥人工智能创新技术的可供性，帮助其适当地融入人类社会的生产生活过程之中，形成“相互尊重、彼此透明、技术交互的创造与实践方式”⑤，体现出预测（anticipating）、自反（reflexivity）、包容（inclusion）、反馈（responsiveness）这四个维度的特性，反映负责任创新的整体过程⑥，强调将责任嵌入创新全过程，对技术和产业创新本身进行伦理反思和价值形塑。不过，由于社会的飞速发展，在此基础上开展反思“负责任停滞”的概念也具有一定合理性，从发展和规避风险的角度来看，负责任创新中“责任”的伦理意蕴，本身就包含“停滞”维度⑦，即“暂停以思考”（stop and think），这也体现了在实践中责任与创新二者之间的动态张力。

（三）大语言模型用户的主体性：核心数智素养的嵌入与培育

生成式人工智能发展带来巨大的经济和社会效应。面对伦理挑战，相关治理主体应当深刻认识到生成式人工智能技术自身在数据、算法方面的限制，同时也应该意识到用户个体的自主性与

① 李韬、周瑞春：《生成式人工智能的社会伦理风险及其治理——基于行动者网络理论的探讨》，《中国特色社会主义研究》2023年第6期。

② 艾瑞咨询研究院：《ChatGPT浪潮下，看中国大语言模型产业发展》，艾瑞网，https://report.iresearch.cn/report_pdf.aspx?id=4166，访问日期：2023年10月22日。

③ 矣晓沅、谢幸：《大模型道德价值观对齐问题剖析》，《计算机研究与发展》2023年第9期。

④ 德里克·德克霍夫：《文化的肌肤：半个世纪的技术变革和文化变迁》（第二版），何道宽译，中国大百科全书出版社，2020年版，第403-404页。

⑤ 阎伟萍、朱春艳：《智能时代理解负责任创新的新思路——读“负责任创新”译丛》，《中国图书评论》2023年第12期。

⑥ 晏萍、张卫、王前：《“负责任创新”的理论与实践述评》，《科学技术哲学研究》2014年第2期。

⑦ 薛桂波、张馨文：《负责任创新“停滞”维度辨析》，《大连理工大学学报（社会科学版）》2024年第2期。

自觉预防风险隐患的积极作用，在全真互联、泛在智能、无限算力的生成式人工智能技术发展浪潮下，增强在技术面前的主动性、意识性和创造性，以及在伦理风险面前的自觉性、自律性和保护性。

从个体媒介素养培育视角出发，这也是从“数字素养”升维成“数智素养”的过程。“数字素养”是指数字获取、制作使用、交互创新、安全保障、伦理道德等一系列素质与能力的集合①，这也是《提升全民数字素养与技能行动纲要2022—2035》的具体要求。面对生成式人工智能带来的伦理道德方面的风险，用户自身较高水平的数字素养，能够帮助其在更为多元丰富的数字实践中精准识别伦理风险，较为审慎进行判断和选择。与此同时，“数智素养”则是数字素养生成式人工智能技术交融时代的素养升维，即数智环境中评估信息资源、运用数字工具、采取数字安全措施的能力，对技术持有合规合理的使用方式和积极的使用态度，在深度上新增了使用ChatGPT、NetObjex、TrustSQL等信息通信技术的使用能力和认知能力②。

四、结语

面向未来，生成式人工智能在资源配置、社会关系、社会秩序和社会结构等社会各方面带来全方位重组。同时，生成式人工智能作为一种新型的技术力量，以其对社会结构的突变性影响，引发不同治理主体对数据安全、技术利维坦、隐私以及认知思维构陷、道德边界模糊等一系列风险问题的关注。在技术高速发展与社会治理框架形成的张力中，更需要找到技术校准的基准点，重新勾勒出大语言模型使用的边界规则与伦理标准，共同促进生成式人工智能技术和产业的健康良性发展。

首先，在宏观层面，需要构建一个全球性的治理框架，以应对生成式人工智能的全球性挑战，在跨国际的治理合作中构建一个安全、公平、可持续的全球性生成式人工智能治理体系，实现广域的数字化联合。其次，在中观层面，需要建立健全制度体系，以规范生成式人工智能的研发、应用和推广的政策法律等“有形边界”，同时进行社会层面的道德规制，找准技术校准点，为保障该技术的健康发展提供更有益的社会环境。最后，在微观层面，需要关注个人的惯习和认知，以“无形之力”搭建起人类与生成式人工智能技术之间的认知边界、信任边界和文化边界，以期在个体层面上提升人们对生成式人工智能的认识和理解。

生成式人工智能是加速锻造新质生产力的“引擎”③，其治理也是一项长期而复杂的任务，需要从法律规制的“有形边界”以及伦理素养的“无形边界”多层面、多视角、多维度出发，多相关方主体协同参与。人工智能技术嵌入并交织在社会文化肌理的当下，面对生成式人工智能技术带来的机遇与挑战，“制衡还是共生”的问题，需要我们在人与技术共生、技术文化与社会协同发展的过程中逐步回答。创新的生成式人工智能治理的行动框架能够变革和培育人类对于技术的认知边界，实现生成式人工智能技术更责任性的创新，以期在未来技术的迭代更新中，生成式人工智能可以在人类社会的历史进程中提升人的自由度，释放并激活其更加强大的社会发展动力。

① 雷晓燕、邵宾：《大模型下人工智能生成内容嵌入数字素养教育研究》，《现代情报》2023年第6期。

② ALA’s Digital Literacy Task Force. Media Literacy for Adults. https://www.ala.org/tools/sites/ala.org.tools/files/content/Media-Lit_Arch_Prog-Guide.pdf, 2023.

③ 于凤霞：《以发展人工智能为重点加速锻造新质生产力》，《中国经济报告》2023年第6期。

Check and Balance or Symbiosis: The Techno-Calibration Points, Invisible Demarcation Line and Action Framework for Governing Generative Artificial Intelligence

YANG Ya SU Fang ZHANG Xueqing

[Abstract] The development of generative artificial intelligence has highlighted the discrepancies between technological acceleration and social governance frameworks. This article considers as its starting point the transformative impact of generative artificial intelligence on social structures as, analyzing prominent issues such as technological risks represented by data security, the technological Leviathan, and privacy security as well as decision-making risks represented by cognitive traps. It seeks benchmarks for the calibration of generative artificial intelligence technology in the context of technological power substitution and social restructuring and takes mid- and micro-level digital literacy cultivation as springboards for action, aiming to provide a governance logic and action framework for the development of generative artificial intelligence beyond tangible boundaries by establishing intangible boundaries and ethical literacy standards. Specifically, at the macro level, it is necessary to construct a global governance framework to address the global challenges of generative artificial intelligence and establish a secure, fair, and sustainable global governance system for generative artificial intelligence through transnational governance cooperation to achieve broad digital collaboration. Second, at the mid-level, it is necessary to establish a sound institutional system to regulate the policy, legal, and other tangible boundaries of the research, application, and promotion of generative artificial intelligence while also conducting moral regulation at the social level, identifying technological calibration points, and providing a more beneficial social environment to safeguard the healthy development of this technology. Finally, at the micro level, attention must be paid to individual habits and cognition to invisibly build cognitive awareness, trust, and cultural boundaries between humans and generative artificial intelligence technology aiming to enhance people's understanding of generative artificial intelligence at the individual level.

[Key words] Artificial Intelligence-generated Content (AIGC); Social Governance; Invisible Demarcation Line; Technical Calibration; Digital Intelligence Literacy

（责任编辑：周瑞春 责任校对：戴瑶）