生成式人工智能责任规制的法律问题研究

袁　曾

（上海大学法学院　上海　200444）

内容提要：生成式人工智能的发展需要法律予以及时规制，以实现技术的发展可受控制。现行人工智能的责任承担规制以算法“可解释”为核心要求，通过算法透明性、隐私保护以及分类分级监管等配套机制构筑了相应治理范式。但在生成式人工智能规模化应用以后，技术的底层逻辑发生了根本性变化，现行架构无法有效调和生产力水平快速提升引致的责任承担等新问题，以“可解释”为中心的责任承担机制需要逐步调整为以“可控制”为中心的人工智能责任规制范式。结合技术能力发展的规律与现实要求，从鼓励与发展生成式人工智能的基点出发，需要基于经济利益与责任承担机制的再考量，重构生成式人工智能的责任规制的核心原则、方式与体系，以期实现规则优势引领发展优势，确保发展“可控制”的人工智能。

关键词：生成式人工智能　ChatGPT　算法治理　算法解释　数字权力

自 1956 年美国达特茅斯会议提出人工智能（Artificial Intelligence，AI）的概念后，人工智能领域一直是学术与实务研究的重点。2023 年年初，以 ChatGPT 程序（Chat Generative Pretrained Transformer）为代表的生成式人工智能（Generative Artificial Intelligence）的规模化应用，再次引

作者简介：袁　曾（1988—），男，汉族，湖北孝昌人，上海大学法学院副研究员，最高人民检察院法治前海研究基地、南方海洋科学与工程广东省实验室（珠海）兼职研究员。

本文为国家社科基金后期资助项目“人工智能的法律人格与未来发展研究”（项目编号：22FFXB026）的阶段性研究成果。

燃了学界对于人工智能话题的关注热情。在生成式人工智能作为颠覆性生产力工具投入实践应用后,其已在认知、责任、侵权等领域对社会形态形成了真实威胁,人工智能可否在人类的预期控制下发展,已经成为必须予以急切关注的重大问题。2023 年 3 月 22 日,美国生命未来研究所发布了一封题为《暂停大型人工智能研究》的公开信,呼吁暂停比生成式人工智能代表性技术 ChatGPT-4 更为强大的人工智能系统训练。[①] 传统人工智能的治理是以算法的“可解释”作为调整与规制相关主体责任的核心规则,若算法无法被解释,则由相应主体承担责任。鉴于人工智能的发展奇点已经到来,传统人工智能的治理模式与责任规制路径已被实践证明无法有效调控技术发展的风险。需要在已有治理经验的基础上,调整生成式人工智能责任承担的核心原则、方式与体系,确保生成式人工智能的发展始终按照预设技术路线为人类所控制,真正实现发展“负责任的人工智能”。

一、现行人工智能责任规制以算法治理为核心

自人工智能技术诞生以来,对于其的有效治理就逐步成为立法者规制科技治理的核心问题,而治理的关键就在于如何解决技术发展过程中的责任承担问题。人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术与应用提供的技术科学。[②] 在生成式人工智能诞生以前,传统人工智能是封闭性的人工智能,通过算法与相对小样本量的数据以解决确定性的问题。[③] 从所处的技术发展阶段与生产力水平出发,当前各国对于传统人工智能的治理基本以算法治理为核心并由此形成了完整的治理规则与责任承担体系。[④]《新一代人工智能伦理规范》第 12 条规定,要在算法设计、实现、应用等环节,提升人工智能的透明性、可解释性、可理解性。人工智能涉及的透明性、隐私保护与可追责性等实际上是以可解释为中心展开的规则架构,其本质是以此确定人工智能致损时的责任承担问题。法律责任的本质是回应自身行为的责任能力,若人工智能作出的行为无法被解释,就不应视作法律意义上的行为。[⑤] 从域内分析,《新一代人工智能发展规划》《个人信息安全规范》《新一代人工智能治

---

① Future of Life Institute. *Pause Giant AI Experiments: An Open Letter*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/, last visited on 2023-05-28.

② See Carl B. FREY, Michael A. Osborne. *The Future of Employment: How Susceptible are Jobs to Computerisation*? Technological Forecasting and Social Change,2017(114): 254-280.

③ See Xinlong Wang, Xiaosong Zhang, etd., *SegGPT: Segmenting Everything In Context*. 6 April 2023, https://arxiv.org/abs/2304.03284.

④ 算法的可解释,是指“使用特定算法进行计算时,人们能够理解并解释该算法所做决策和结果的程度。即在算法运行过程中,人类是否能够理解算法的决策、运行过程和结果,而不仅仅是得到准确的结果”。See Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a “Right to an Explanation” is Probably Not the Remedy You are Looking For*, 16 Duke L. & Tech. Rev. 18(2017).

⑤ 欧盟 2020 年 10 月发布的《关于人工智能道德原则的立法倡议》提出,人工智能作出侵权或损害行为,是算法在消化吸收数据后由算法决策的错误所致,这种失误既可能来自系统所提供的信息或决定,也可能来自所训练的数据集,还可能来自外部干扰或者新型介质。但错误的来源或原因无法确定时,人工智能的不透明性、规模效应及无可避免的歧视等特性,就很可能引致不同的犯罪责任后果。参见刘艳红:《人工智能的可解释性与 AI 的法律责任问题研究》,载《法制与社会发展》2022 年第 1 期。

理原则》《关于规范和加强人工智能司法应用的意见》《算法推荐管理规定》《互联网信息服务深度合成管理规定》等大量的政策法规文件均提出了“安全可控”的人工智能治理目标，明确要求通过算法透明、算法设计、算法输出等具体规范建立公开透明的责任体系，实现算法的全流程监管。[⑥] 从域外分析，欧盟的《人工智能法案》《通用数据保护条例》《算法问责及透明度监管框架》《可信人工智能伦理指南》以及美国的《电子记录系统和个人隐私》《公平信贷报告法》，均规定了技术持有者对于算法的义务要求。[⑦]

考虑算法技术持有者的实际地位，当前算法的责任规制体系以“可解释”为核心发展建构，以算法持有者责任为基础，通过算法备案、分层分级分类治理等配套措施与规制要求，确定相应的责任义务。[⑧] 在实际的侵权事件中，若算法持有者无法对决策的目的、原因与过程提供清晰的解释说明，则应承担相应的不利后果。传统人工智能下，算法依靠小样本数据量下的模型计算实现产出，算法的持有者具有单方优势地位，受众仅能被动接受算法的控制，在出现系统性风险或具体的错误时，通过寻求算法的行为逻辑以确定持有者的责任路径具备相应的逻辑架构与现实可能。例如，在 Bauserman v. Unemployment Insurance Agency 案中，美国密歇根州政府使用的第三方算法系统错误地认定部分居民存在骗取社会福利的行为，从而遭到该洲居民的集体诉讼。[⑨] 基于算法“可解释”形成的责任治理架构，在传统人工智能时代有其存在的实际价值与意义。[⑩]

二、现行规制体系无法有效解决生成式人工智能的责任问题

现行人工智能责任规制建构的核心规则与架构，基本可以应对传统人工智能时代的现实问题，但在生成式人工智能正式投入商用以后，传统以算法可解释为规制核心的责任承担机制已无法应对新生产力工具带来的颠覆性变化。生成式人工智能以大型语言模型为基础，通过深度学习自互联网诞生以来全域范围内的海量数据，演绎归纳形成符合逻辑的输出内容，其利用数据之广泛、生产内容之准确、算法迭代之迅猛、可利用范围之宽广，已非传统人工智能可以匹敌。如何通过责任规则与体系的建构，确保生成式人工智能始终在人类的控制下负责任地发展，是学界研究的焦点与舆论关注的热点。

（一）技术能力突破了传统法律规制范围

一是生成式人工智能改变了传统人工智能的认知局限。传统人工智能是以深度学习算法为技术逻辑，生成式人工智能在此基础上应用了更为广泛的训练样本，即更为庞大的基础

---

⑥ 参见许可：《驯服算法：算法治理的历史展开与当代体系》，载《华东政法大学学报》2022 年第 1 期。

⑦ 美国联邦贸易委员会（FTC）2020 年发布的《人工智能和算法运用》与 2021 年发布的《公司运用人工智能：以真实、公正、平等为目标》两份解释性规则，指出应用算法者应采取保持透明、解释决定、决定公平、数据与模型可靠、可问责五类措施，要求算法持有者需满足相应的法律要求。参见马海群、蔡庆平等：《美国数据与算法安全治理：进路、特征与启示》，载《信息资源管理学报》2023 年第 1 期。

⑧ See Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 Admin. L. Rev. 1 (2019).

⑨ Bauserman v. Unemployment Insurance Agency, No. 156389 (April. 5, 2019).

⑩ 参见郭春镇、黄思晗：《刑事司法人工智能信任及其构建》，载《吉林大学社会科学学报》2023 年第 2 期。

数据。[11] 生成式人工智能是由“算法”转向为“算法+算力”的技术建构逻辑,相较传统人工智能的责任承担机制仅对“算法”主体加以大量的治理义务,显然无法覆盖“算力”领域的规制要求。例如,在数据的抓取过程中,即便算法的设定中已经确定了避免抓取权利人隐私或涉及知识产权的作品,但由于样本数量的过于庞大,算法可解释也无法回避侵权的必然结果,由此无法解决责任分配与承担的问题。

二是生成式人工智能具备了推理能力等意思表示特征。传统人工智能基于算法形成的结论,可以回溯算法的研发过程与运行过程,其不具备意思表示的行为架构与推理能力,而常识与推理的结合才能够形成人类智能的特征。但生成式人工智能已经突破了能力局限,其已可以做出超出算法预设的可期结果。若向 ChatGPT-4 展示第二天的雨雪预报图片并询问出行建议,ChatGPT-4 会准确凭借天气状况给出应乘坐公共交通工具的建议。在此维度上,生成式人工智能已经实现了自身意思表示,直接导致了算法“可解释性”责任架构下的逻辑链条割裂。

三是生成式人工智能已经具备了产生相应后果的行为特征。早在 1972 年,美国法院就在 State Farm Mutual Automobile Insurance Company v. Bockhorst 案中承认了智能系统在决策过程中所做的决定有效,应由被代理人承担智能系统的行为后果。[12] 在生成式人工智能规模化商用后,其行为能力得到了进一步的强化。若有主体利用生成式人工智能,广泛注册虚拟社交媒体账号并设置隐藏的特定输出内容,极有可能会对其他使用者产生负面影响,此种情形下即便算法原理与运行过程可以被解释,也无法解决实践中产生的负外部性问题,由何种主体承担责任尚属法律规制的模糊地带。

(二)现行规制架构无法涵盖现实治理要求

当前人工智能的治理是以规制算法的系统性法律为基础框架,对算法的设计、使用和数据种类提出明确的要求与标准,以此确定责任承担的主体与具体履行义务的方式,在一定时期内确保了我国在治理对象、链条与工具上的法律制度优势,但技术迭代的迅猛发展已超过了原有治理框架的能力范围。[13] 由于传统人工智能运用的核心技术模型仅凭人类直觉难以理解,将算法决策转化为可解释的推断过程,可以使用户理解和相信算法决策,由此确定算法的技术主体承担责任具备现实合理性。而生成式人工智能使用的数据样本数量远超过传统人工智能,即便算法逻辑可以被解释,当算力与数据的结合实现突破以后,人工智能的技术能力也随时可能出现“涌现”(Emergence),突破算法的预先设置。[14] 就此维度而言,探究算法可解释的实践意义在生成式人工智能时代已经偏离了现行责任承担规制最初设定的目的基础。由于包含操作者等多种主体介入的复杂性,当技术已经实现从客观事实形成主观判断的能力时,即便对于底层

---

⑪ 参见姬蕾蕾:《私密信息界定的司法困境及其破解方向》,载《上海大学学报(社会科学版)》2022 年第 6 期。

⑫ State Farm Mutual Automobile Insurance Company v. Bockhorst, 453 F. 2d 533(10th Cir. 1972).

⑬ 参见张凌寒:《深度合成治理的逻辑更新与体系迭代——ChatGPT 等生成型人工智能治理的中国路径》,载《法律科学(西北政法大学学报)》2023 年第 3 期。

⑭ 参见[以色列]尤瓦尔·赫拉利:《人类简史:从动物到上帝》,林俊宏译,中信出版社 2014 年版,第 400 页。

算法的解释程度再高，也无法准确预判生成式人工智能输出结果的多样性，算法“可解释”与结果“可信任”之间的唯一关系链条已被发散，责任的承担变得极为复杂。这类似于刑法中要求罪犯解释犯罪动机，但犯罪动机的解释不清或无法解释并不妨碍犯罪的构成。特别是生成式人工智能使用的是全域数据，算法本身无法决定输入变量的误差与基准，而变量的超预期变化，使得“可解释”的可能性与重要性均大幅度降低，可解释的责任规制逻辑已失去了原有的信任意义。因此，需要从全局治理的维度出发，及时修正人工智能责任规制的模式与机制。

（三）现行规制架构无法妥善处理义务承担问题

一是现行架构不符合生成式人工智能乃至未来通用人工智能发展的责任建构逻辑。包括《通用数据保护条例》在内的法规明确要求算法“可解释”的逻辑，在于确定人工智能致损时的损害原因，以确定责任主体。[15] 此种机制，实际上将责任完全限制在算法的研发者与提供者身上。2023 年 4 月 11 日，国家网信办发布《生成式人工智能服务管理办法（征求意见稿）》，以促进生成式人工智能健康发展和规范应用，其主要内容以规制智能产品或服务的提供者为主要抓手，仍属于传统人工智能治理的责任模式。生成式人工智能的技术应用中，增加了数据所有者、算力供应商、模型设计者等多圈层的多样化主体，仍对生成式人工智能的提供者分配最重的责任义务具有天然的不公平性，也无法处理社会发展中使用新科技而引发的新生问题。

二是无法解决生成式人工智能自身的责任问题。“生命最大的悲哀，莫过于科学汇聚知识的速度快于社会汇聚智慧的速度”。[16] 生成式人工智能已具备了类人的思维形态与一定的行为能力，使得现有法律规制模型无法适应技术的快速变化。以民法调整的领域为例，生成式人工智能的强大算力与大模型的技术加持使得无人驾驶真正成为可能的现实，但当无人驾驶汽车成为侵权主体时，如其自行行驶撞伤前车突然意外掉落的乘客时，应由何种主体负责，是现有法律规定的空白。在此情形下，即便要求无人驾驶技术的提供者承担“可解释”的义务，其与具体侵权事件的因果关系厘清、责任承担也难以存在直接的联系。[17] 再以刑法调整的领域为例，由于生成式人工智能可能产生无法预期的输出内容，可能造成对于特定群体的歧视与侵害，规模化训练的大语言模型很可能造成严重后果。[18] 若仅要求算法的设计过程与逻辑可以被解释，却未将生成式人工智能自身作为特殊主体予以相应的修正或纠错，则显然脱离了责任能力的现实。

三是对单一主体施加过重的义务，无法满足基本经济规律。在传统人工智能阶段，算法“可解释”的要求较易满足，在进行备案登记后较易管理，从治理成本与责任分担上分析是合理

---

⑮ 参见刘友华、李扬帆：《短视频平台强制性版权过滤义务的质疑与责任规则的优化》，载《法学杂志》2023 年第 3 期。

⑯ ［美］迈克斯·泰格马克：《生命 3.0：人工智能时代人类的进化与重生》，汪婕舒译，浙江教育出版社 2018 年版，第 418 页。

⑰ 参见袁曾：《无人驾驶汽车侵权责任的链式分配机制——以算法应用为切入点》，载《东方法学》2019 年第 5 期。

⑱ 参见刘艳红：《Web 3.0 时代网络犯罪的代际特征及刑法应对》，载《环球法律评论》2020 年第 5 期。

的。[19] 但由于运行机理与建构逻辑发生了重大变化，生成式人工智能在操作者的指示下输出内容，其设计的逻辑与产出的结果间，很可能超出算法提供者或研发者的设计初衷。例如，若操作者利用 ChatGPT 实施诱导型犯罪，在此情况下要求 OpenAI 公司承担算法解释的义务并承担相应责任显然过于偏颇，不利于保护技术的创新发展。

（四）现行规制架构无法有效应对监管难题

在进入以区块链、元宇宙等新科技为代表的全新社会生产力阶段后，人类文明的形态呈现出更为明显的数字化倾向。由于生成式人工智能的研发投入巨大、技术鸿沟明显、颠覆能力极强，在其进入规模化应用以后，此种数字权力的中心化垄断趋势不断强化，若继续适用以“可解释”为核心的算法备案或算法监管体系，对生成式人工智能的监管无异于盲人摸象，无法触及其底层逻辑与控制核心。生成式人工智能的信息处理能力与人类通过视觉获取信息而产生意识反应的复杂机制相近，但模拟人类的认知过程与人类的认知过程存在相同的弊病，即无法实现每次运算的结果均在同一可控的范围内。[20] 由于数据输入、数据库样本差异等原因导致的变量大量存在，监管者也无法预料生成式人工智能在操作者的使用下会输出何种内容，算法可解释与监管有效性间的逻辑缺少实践联系。随着生成式人工智能自身自我学习与数据训练能力的快速强化，绝大多数数据管理机构与普通民众在不知情的情形下极有可能已被爬取与使用相关数据和信息，但监管机构并无与之匹配的能力工具对其加以监管。人类代际传承繁衍依靠的是听说读写文字的基本技能，从应用维度上分析，生成式人工智能具备了对于自然语言的理解能力，而受众又依赖于生成式人工智能产出的交互内容，由于技术壁垒与资本屏障，生成式人工智能的拥有者可以在监管者毫不知情的情况下输出颠覆国家政权、消灭文明多样性的潜在价值内容。在此深刻历史背景下，现行规制架构仅要求算法“可解释”与分类分级监管，但又缺少监管能力与规则框架，使得生成式人工智能的监管流于形式。

三、责任规制需要以“可控制”为核心

马克思指出，不是人为法律而存在，而是法律为人而存在。[21] 正视技术迭代发展的现实，对生成式人工智能的治理需要由以“可解释”为核心逐步转向以“可控制”为核心的责任治理架构。“可解释”的体系性要求在传统人工智能时代可以适应其技术逻辑与科技能力，通过穿透人工智能黑箱，确定算法研发者与提供者的责任，在一定时期内发挥了良好的规制作用。但继续固守“可解释”的治理内核，已无法适应生产力水平提升对于社会整体关系变革与未来发展的治理要求。我国人工智能产业在应用维度拥有较强的先发优势，但大部分集中在价值链条较低的环节，与发达国家有较为明显的差距。[22] 通过法律调整责任的合理分配以实现规则治

---

⑲ 参见邱泽奇：《算法治理的技术迷思与行动选择》，载《人民论坛·学术前沿》2022 年第 10 期。

⑳ 参见刘伟：《人机混合智能：新一代智能系统的发展趋势》，载《上海师范大学学报（哲学社会科学版）》2023 年第 1 期。

㉑ 参见《马克思恩格斯全集》（第 3 卷），人民出版社 2002 年版，第 40 页。

㉒ 参见苏玺鉴、胡安俊：《人工智能的产业与区域渗透：态势、动力、模式与挑战》，载《经济学家》2023 年第 2 期。

理,可以获得巨大的理论先发优势。[23] 发展可信向善的人工智能是各国已形成的基本共识,以此为基点,根据生产工具的技术特征与发展规律适时调整治理原则,从经济基础的底层内涵出发建构完整的责任承担机制,根据数字社会变迁的发展方向进行相应的法哲学方法调整,已经成为现实而紧迫的法学命题。控制,是指为了确保组织内各项计划按规定去完成而进行的监督和纠偏的过程。[24] 生成式人工智能"可控制"需要在给定目标的基础上,就技术发展的逻辑,设定治理目标与范式,从基础规则出发即介入控制,从起点至结果全流程介入干预人工智能的治理。生成式人工智能"可控制"是一项系统性工程,其机理是将技术发展的核心价值、底层逻辑、路径规制、关键救济等全过程始终掌握在人类手中。"可控制"的要求并非颠覆原有"可解释"的要求与有价值经验,而是在"可解释"的基础上纵深发展人工智能可控的治理链条,延展控制的范围、内涵与方法。其核心是调整人工智能治理的责任规制重心,扩大"可控制"的责任调整范围。

在人工智能的底层逻辑中事先构建完善的底层动力,是未来进行干预的技术基础和社会保障。[25] 原有"可解释"的控制基础来源于对算法提供者的限制,通过备案、分层分级治理等手段实现监管目的,可控制的主体少、流程短、技术能力弱,需要予以扩张性修正。就技术本身而言,生成式人工智能分为算力、数据、算法这三大类要素,就实际应用而言,生成式人工智能又涉及技术研发者(模型设计者)、技术持有者(数字科技资本)、数据供应者(数据库)、数据所有者(数据主体)、技术使用者(人工智能操作者)、技术监管者(生态治理主体)这六大类主体。以算法"可解释"为核心的责任规制架构,可以涵盖的调整范围仅仅是生成式人工智能治理中的部分领域,需要按照"可控制"的基础目标与终极价值,系统性重构责任规制范围,将三大要素与六大主体以技术的有效利用为主线,全部纳入控制的范围统一调整,而非仅仅要求算法的提供者履行解释及相应配套义务,有关"可控制"责任体系的具体建构在下一章详述。

在确定"可控制"的人工智能责任治理向度后,关键是扩展人工智能的治理方法。其一,有效利用新技术工具治理新科技,即发挥人工智能自主性,扩展适用"以人工智能治理人工智能"。以指令式绘图软件 Mid-Journey 为例,由于操作群体的基数庞大,其每小时在全球范围内生成的各类绘画作品数以百万计,若对其予以传统人工审核,显然无法覆盖如此规模庞大的样本,对如此海量的数据生成进行全域监管也无法覆盖监管成本。[26] 借助强大的人工智能技术对其他技术进行监管就成了符合治理综合效益与新科技特点的选择,根据"可控制"的原则,要求程序母公司在运算中加入可识别的底层水印标记,若生成式人工智能输出的内容缺乏源水印标识,则需要生成式人工智能的提供者承担管控不能的不利后果,而非要求其解释该作品是根

㉓ 参见唐林垚:《"元宇宙"的规制理论构建及中国方案》,载《上海大学学报(社会科学版)》2022 年第5 期。

㉔ 参见胡梦云、黄何婷:《控制论视域下的粉丝侵权行为归责问题初探》,载《财经理论与实践》2022 年第4 期。

㉕ 参见喻国明:《ChatGPT 浪潮下的传播革命与媒介生态重构》,载《探索与争鸣》2023 年第 3 期。

㉖ See Baidu, Human creators stand to benefit as AI rewrites the rules of content creation, MIT Technology Review, https://www. technologyreview. com/2022/11/29/1063736/human-creators-stand-to-benefit-as-ai-rewrites-the-rules-of-content-creation/, last visited on 2023-05-26.

据何种逻辑运算产生。其二，适应技术发展的逻辑与能力，调整规范生成式人工智能的责任标准。我国《新一代人工智能发展规划》《国家新一代人工智能标准体系建设指南》均强调了人工智能标准体系的建设，技术标准则承载了规则要求，使规则中的权利和义务更加具体。[27] 但是传统的以“可解释”为中心的算法标准，如当前主要适用的侵权风险责任、歧视性风险等标准，均过于强调算法研发者与提供者的责任，无法回应生成式人工智能时代其他主体（如操作者）的责任标准要求。[28] 由于大部分生成式人工智能需要在操作者的指令下输出相应内容，缺乏对于输入的标准限定将不利于对其的深度治理。若通过限制生成式人工智能的使用，从短期看可以减少意识形态的干扰，但长期看无法顺应生产力的发展变革，因此通过操作者使用端的标准限制以实现“可控制”的应用就变得颇为关键，如出台国家级标准明确不得使用生成式人工智能参加的考试类型与范围、在中小学设置合理使用生成式人工智能技术的课程标准等，以尽可能避免技术马太效应。其三，修正生成式人工智能的技术评价导向，系统性建构通用人工智能发展方向下的人工智能评价体系。在“可解释”的评价体系下，算法在解释程度上的优劣与责任承担直接挂钩，要求算法实现的目标尽可能地精准、统一。在生成式人工智能加入操作者等重要主体的介入后，完全统一的精准输出受到输入条件的限制已经很难实现，因此引入新的评价指标才能更好地契合技术的发展。“可控制”不仅是责任规制原则的重心调整，而且是通过系列方法工具、应用标准、评价体系的系统性变化，以实现技术发展责任的合理分配，由此实现生成式人工智能的整体可控。

四、完善以“经济利益—责任承担”为逻辑的责任规制基础

马克思在1859年《政治经济学批判》的序言中指出，经济基础决定上层建筑。算法“可解释”导向下的责任规制架构将责任过分集中于技术的研发者与持有者身上，造成了单一主体的责任义务过大，又缺少相应的制度构建保障其经济利益。当前舆论对于数据平台企业呈现出过于苛责的倾向，“陷在算法里的骑手”“算法即剥削”等观点反映了从业者与普通民众对于垄断型技术平台的忧虑，这类看法来源于社会实践，有其相应的合理性。但从客观上分析，其深层次矛盾实际是技术研发者、持有者等科技资本在实际收益与责任承担间的冲突，更反映的是规制模式与分配制度未能适应生产力发展的变化要求。就改变责任承担范式的价值维度而言，生成式人工智能的治理背后体现的是数字经济发展过程中法律规制调整技术发展、社会变迁乃至标准制定权、引导权的规则之争，若在核心理论上予以突破，则极有可能在实践的正回馈支撑下极大地改善营商环境，通过规则的领先最终实现技术、资本的繁荣与国力的竞争优势。[29]

㉗ 参见关保英：《论大数据时代的行政法精神》，载《上海政法学院学报》2023年第1期。

㉘ 例如，全国信息安全标准化技术委员会于2022年5月发布的《网络安全标准实践指南——人工智能伦理安全风险防范指引》中提出的标准。

㉙ 参见李世刚、包丁裕睿：《大型数字平台规制的新方向：特别化、前置化、动态化——欧盟〈数字市场法（草案）〉解析》，载《法学杂志》2021年第9期。

（一）建构符合经济规律的新型责任分配机制

笔者在2019年针对学界当时热议的无人驾驶汽车对于责任认定的冲击，提出了链式责任分配机制，即抛弃原有将侵权责任集中在无人驾驶汽车生产者或研发者的理论设想，将侵权责任合理分配于汽车的研发者、生成者、使用者、监管者与保险人等主体之上，通过因果关系的查证与相应配套机制的完善，在保证责任有主体承担的基础上，保证技术研发的投入与发展。[30] 该链式责任分配机制在生成式人工智能进入应用后，具备相同的法学、经济学理论基础与更加现实的技术可能，基于经济利益基础的考量匹配相应的责任规制，是“可控制”治理体系下的责任规制框架范式。生成式人工智能至少涉及技术研发者、技术持有者、数据供应者、数据所有者、技术使用者、技术监管者这六大类主体，以技术的使用为主线，将经济利益与责任承担深度绑定，形成完整的链式责任分配机制，确保责任均由相应主体承担且在一般情形下单一主体的责任不至于过重，实现生成式人工智能的总体可控发展。

（二）通过结构化规制实现整体责任清晰

生成式人工智能链式责任分配包含以下几个层级的内容，通过相应主体的分段分层可控，实现生成式人工智能发展的总体可控。其一，技术研发者承担研发责任，履行明确的负面清单义务与人机伦理要求，在违法研发明确禁止的技术时承担严格责任（如人工智能操作的胎儿基因编程技术）或无过错责任，在其他情形下承担过错责任，并可自由选择是否投保相关保险以应对科技风险。其二，技术持有者承担运营主体责任，根据“可控制”要求下新的技术标准与评价标准履行义务，履行数据保护、内容管控、溯源标记等明确义务，通过投保强制责任保险等规避系统性风险。除利用生成式人工智能颠覆国家政权、输出暴力色情内容等违法行为以外，其仅应承担适度的责任义务，鼓励其投资新型技术、扩展算力规模并保护其技术收益。其三，由于生成式人工智能需要使用海量的数据进行运算，因此数据供应者与数据所有者的责任与权益需在经济基础的考量上重新概念化。数据是新科技时代最为重要的资源，目前我国学者对于数据确权的研究理论成果丰富，但是由于生成式人工智能时代数据均为规模化共享、使用和流动，对于数据的单独确权极为繁琐且意义不大。欧盟及美国均未从国家立法的层级确认数据财产权地位，我国《民法典》第127条并未给数据确权，但学界对数据确权进行了大量尝试，可这并不符合数据作为基本生产要素的特点。[31] 顺应数据规模化利用的潮流，可以重新定义数据的定价机制与数据供应、使用的责任，监管者可以设立独立的生成式人工智能数据收益基金，无论使用抑或交换数据产生的部分收益均统一纳入该基金管理，当产生非因数据供应者或

㉚ 例如，当前对于“美团”公司等外卖平台的诟病，主要存在于其对骑手的送餐时间短、送餐单价低等方面，该企业经过多次调整后形成了优化后的配送规则与算法，但仍然不尽如人意。根据美团公司公开的2022年报，其当年营业收入为2199.55亿元，仍净亏损66.86亿元。规模庞大的数据平台企业需要大量资金投入与维护运营成本，单纯增加骑手派送费只会增加企业成本，但又无法回馈到企业收益上，而美团公司底层庞大的可供生成式人工智能所使用的数据库、模型等又未有明确的标准与方法转化为营收，最终的结果只能是勉强维持现状。参见袁曾：《无人驾驶汽车侵权责任的链式分配机制——以算法应用为切入点》，载《东方法学》2019年第5期，第28页。

㉛ 参见周汉华：《论平台经济反垄断与监管的二元分治》，载《中国法学》2023年第1期。

所有者原因引发的泄露或侵权时，由该基金承担相应的风险。数据的规模化应用的风险责任与定价要素是极为复杂的问题，需要在技术发展的过程中不断调整以实现平衡。其四，由于生成式人工智能在操作者的介入下可以产生无法控制的结果，因此对于操作者的义务也需要有明确的负面清单限定与禁止性规范，当操作者利用生成式人工智能实施违法犯罪并可以溯源时，应由操作者自身承担相应的责任。其五，生成式人工智能的监管者负责搭建适应生成式人工智能可控发展的环境生态并承担最终监管责任。钱学森在1981年曾前瞻性地提出“人—机—环境系统工程”理论，运用系统科学理论和系统工程方法，通过正确处理人、机、环境的要素关系，寻找最优组合以实现系统“安全、高效、经济”。生成式人工智能首先是人机环境系统，环境是做事的平台，如元宇宙等新型环境空间。[32] 监管者在新科技时代需要按照可控制的目标逐步调整监管方法与理念，以趋向技术治理的方式搭建符合人工智能发展方向的底层环境生态。例如，在中国大语言模型的发展生态上，优先支持训练以中文为主体的大模型，注重中文数据质量的构建与结构化配置。当生成式人工智能具备了不可控风险时，由监管者承担规制、打击乃至取缔等不同层级的控制义务。

（三）权利与责任适度剥离

由于生成式人工智能具备了类人化的认知推理能力与行为能力，导致现行法律体系难以应对技术能级提升后的现实问题。以知识产权为例，生成式人工智能输出的内容是否可以成为作品即牵涉众多复杂问题，若不承认其输出的内容为作品，其又具备现行著作权法对作品的一般特征，“创造力与创新能力除了所属主体存在着不同之外，在其他的规定性方面并没有太大差别”。[33] 若承认其输出的内容为作品，则应由何种主体享有财产权？[34] 再以刑事犯罪为例，网络与数据的特性使得罪犯可以超越物理空间在多个国家或地区同时利用人工智能实施犯罪并隐匿身份行踪、销毁或隐藏关键证据，对传统以地域为核心的管辖提出了严峻挑战。[35] 在穷尽技术手段也无法查证犯罪主体的情形下，如何归责、追责？再如，当前在视频网站B站上流行的AI歌星形象，若被用于深度伪造生成色情影片，则如何厘清数据平台与生成式人工智能持有者的责任关系？面临诸多棘手现实问题，可以适时考虑引入新的权责模式，赋予生成式人工智能特殊的法律地位，将权利与责任适度剥离，以使得生成式人工智能的侵权损害赔偿责任最终可以有主体承担，避免陷入无人担责的“公地悲剧”。基于生成式人工智能的现实行为能力，可以赋予其一定的权利能力，使其拥有相对独立的法律地位，因此享有部分财产性权利。但这种财产性权利又是特殊的，需要在监管的框架下独立归集，以便在出现侵权事件时可责令侵权者承担相应的损害赔偿责任。例如，承认生成式人工智能输出的内容为作品，其财产性权利由人工智能所

㉜ 参见赵精武：《“元宇宙”安全风险的法律规制路径：从假想式规制到过程风险预防》，载《上海大学学报（社会科学版）》2022年第5期。

㉝ 参见高新民：《创造力的计算建模、机器实现及其认知哲学意义》，载《上海师范大学学报（哲学社会科学版）》2022年第1期。

㉞ 参见黄玉烨、司马航：《孳息视角下人工智能生成作品的权利归属》，载《河南师范大学学报（哲学社会科学版）》2018年第4期。

㉟ 参见裴炜：《网络空间刑事司法域外管辖权的数字化转型》，载《法学杂志》2022年第4期。

有,其他主体在使用该作品时应当支付相应的微量费用,这些费用累积进入生成式人工智能提供者按监管要求设定的独立账户内,用于生成式人工智能在抓取、使用数据或他人作品时侵权所造成的损失,从经济的角度解决现实矛盾。就法律的发展过程而言,立法者对于主体的法律地位或称法律人格的调整始终处于一种变化但总体呈现扩展的态势,法律人格的扩张与生产力水平的提升与生产关系的变化密切相关,女性、婴儿等享有的权利能力均是在人类认知达到一定水平之后方才形成的。就生成式人工智能的权利能力、行为能力与责任能力而言,对其适用特殊的法律规则并给予特定的法律地位,类似于大航海时代拟制公司法人的做法,体现的是经济社会发展需求与法律技术进步间的正相关关系,最终解决的仍然是人类社会的发展问题。[36]

五、结论——数字社会背景下法哲学方法的调整

生成式人工智能是通用人工智能的技术起点,生成式人工智能的诞生对于社会而言是一项系统性的演变。[37] ChatGPT 等智能的实际应用已经打开了人类正式迈向智能数字社会的大门,对其研究探讨的实际意义已经超越了元宇宙、Web3.0 等仍较为遥远的概念,其对于社会运行形态与生产关系的改变是颠覆性的。[38] 当数字技术已初步具有类人认知推理与行为能力时,传统社会的结构与逻辑必然会受到挑战,在数字社会的维度下,应当适用何种法哲学以适应并调整生产力的发展水平,需要进一步廓清和体系化。[39] 有学者提出,大数据时代下任何制度建构均须依赖于相应的数据,法律趋于数字构型是新的趋势,要用新的数字技术构造法律体系和评判法治过程。[40] 但也有学者认为,法律的生命向来不是逻辑而是经验,法律所承载的是民族演化的历史叙事而不只是数学般的定理和定论。[41] 路德维希·维特根斯坦在《逻辑哲学论》中指出,“世界是事实的总和而非事物的总和”。[42] 在不同的法哲学观导向下,必然形成不同的治理方向与规则体系。

生成式人工智能事实上已在众多领域替代了人类工作,在促进生产力水平提高的同时,也带来了权益侵害、隐蔽犯罪、认知混乱等客观风险。正视实际,发展可信向善的人工智能,需要按照生产力发展的规律适时调整生产关系,其关键在于根据“可控制”的责任规制总体要求,给予生成式人工智能各参与方以恰当的责任机制与归责路径,及时监督与修正预期外的重大风险,在技术进步与治理安全的综合目标下形成准确的价值预期,在有效规避系统性风险的基础上保护创新,促进技术的整体可控发展。法学是社会科学,其存在的目的与意义是为解决社会

㊱ 参见田宏杰:《法律与道德:正义的法哲学及其发展——以瑞特纳帕拉法学思想研究为核心》,载《法学杂志》2022 年第 1 期。

㊲ 参见肖峰:《生成式人工智能与数字劳动的相互关联——以 ChatGPT 为例》,载《学术界》2023 年第 4 期。

㊳ 参见王少:《ChatGPT 介入思想政治教育的技术线路,安全风险及防范》,载《深圳大学学报(人文社科版)》2023 年第 2 期。

㊴ 参见马长山:《数字何以生成法理?》,载《数字法治》2023 年第 2 期。

㊵ 参见关保英:《论大数据时代的行政法精神》,载《上海政法学院学报》2023 年第 1 期。

㊶ 参见关保英:《行政法代际问题研究》,载《法学杂志》2022 年第 5 期。

㊷ [奥]路德维希·维特根斯坦:《逻辑哲学论》,贺绍甲译,商务印书馆 2020 年版,第 22 页。

实践问题,并促进社会的稳定发展。[43] 从数字社会视角分析,以目标导向为评价基准,基于技术发展方向与经济社会利益基础,逐步调整规范“可控制”的人工智能以实现其可信向善的发展方向,很可能是立法者基于实用主义与技术治理的应然选择。

Research on Legal Issues of Liability Regulation of Generative AI

Yuan Zeng

Abstract: The need for generative artificial intelligence requires timely regulation by law so that the development of the technology can be controlled. The current AI liability regulation takes the algorithm “explainable” as the core requirement, and constructs the corresponding governance paradigm through the algorithmic transparency, privacy protection and classification supervision. However, after the large-scale application of generative Al, the underlying logic of the technology has undergone fundamental changes. The current structure) cannot effectively reconcile the practical problems caused by the rapid increase of productivity level, and the governance requirements centering on “explainable” need to be gradually adjusted to the Al governance paradigm centering on “controllable”. Combined with the law of development of technological capabilities and practical requirements, from the basic point of encouraging and developing generative Al, it is necessary to reconsider the governance principles, methods and systems of generative Al based on economic interests and responsibility assumption mechanism with a view to realizing the advantage of rules leading to the advantage of development and ensuring the development of “controllable” Al.

Keywords: Generative Al; ChatGPT; algorithm governance; explanation of algorithm; digital power

(责任编辑:刘宇琼)

㊸ 参见冯玉军:《高质量立法为良法善治奠基——兼论〈立法法〉再次修改的理由和要点》,载《法学杂志》2022 年第 6 期。