

大数据·人工智能·区块链研究（二十三）

# 从 ChatGPT 审视负责任创新： 阻碍因素与转型方向

胡泳 朱政德

（北京大学新闻与传播学院，北京 100871）

**摘要：**ChatGPT 的勃兴再次激化了技术创新加速和社会适应水平滞后两者间的尖锐矛盾，使人类被迫面临前所未有的经济、政治、文化和环境风险，进而使有关负责任创新的呼声愈演愈烈。但作为技术伦理准则的负责任创新沿袭了熊彼特范式下基于竞争逻辑的创新，其设计的 ARDR 模型一方面从未在根本上调控创新较之于责任的支配性优势，另一方面把负责任简化为线性流程，导致负责任看似获得与创新并列的地位，实则在创新蕴含的人际竞争、公司竞争、国际竞争面前屡屡沦为点缀。唯有在后熊彼特范式下依托基于共生逻辑的创新保卫负责任，才能建构起真正具备约束力的伦理准则，削弱不负责任创新赖以扩散的社会基础，使通过负责任停滞促成制度跟进变为可能。实现上述目标不仅要求各国完善内部治理，而且需要负责任的大国制定跨国人工智能治理框架，使人工智能的发展对全人类有益。

**关键词：**ChatGPT；负责任创新；技术伦理建设；竞争逻辑；共生逻辑

**中图分类号：**G206 **文献标识码：**A **文章编号：**1005-9245（2024）01-0139-12

## 一、问题的提出

20 世纪 70 年代末，在双陆棋程序 BKG9.8 代表智能体首次击败此领域人类世界冠军<sup>①</sup>前夕，美国技术哲学家兰登·温纳（Langdon Winner）指出：“技术系统的发展总是挑战个体和社会的适应水平，技术创新加速导致预测特定创新的影响范围愈发重要也愈发困难。社会—技术系统日渐复杂，

社会生活的最基本行动——规划、设计、功能协调都陷入窘境。”<sup>②</sup>随后不到半个世纪，全球人工智能市场凭借近 52% 的年复合增长率臻于鼎盛<sup>③</sup>，DeepBlue、AlphaGo、AlphaStar、ProjectDebater、DeepNash 等使智能体战胜人类的记录遍布博弈、辩论、电子竞技等诸多领域，当 ChatGPT 以全知全能（Almighty）的姿态攻陷内容生产这一最能凸显人类才智的领域时<sup>④</sup>，人工智能全面凌驾人类的

收稿日期：2023-05-30

作者简介：胡泳，北京大学新闻与传播学院教授、博士生导师；朱政德，北京大学新闻与传播学院博士研究生。

① M.V.Zelkowitz, et al. Fortieth Anniversary Volume: Advancing into the 21st Century, Advances in Computers, San Diego, CA: Academic Press, 2000: 211.

② L.Winner. Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought, Cambridge, MA: MIT Press, 1977: 3.

③ K.Niemiec, N.Katarzyna, K.Damian. Intelligent Tools for Improving Quality and Efficiency in Production Environment, Journal of Physics: Conference Series, 2022 (1).

④ M.Cheong. ChatGPT, the Almighty AI, is a Neoliberal College Graduate, <https://www.rt.com/news/571363-chatgpt-woke-ai-liberal/>.

“奇点”被认为触手可及<sup>①</sup>。

作为一款基于人工神经网络、由 45TB 语料数据(包含 8000 亿个单词和 1750 亿个参数)训练而成的大语言模型,ChatGPT 采取独特的半监督式深度学习法,在回答人类提问时能够逼真模仿人类语言风格、灵活揣测后文信息、精准贴合提问条件,堪称前代智能体的集大成者<sup>②</sup>。ChatGPT 已攻克“美国高考”SAT、美国执业医师资格考试、宾夕法尼亚大学沃顿商学院 MBA 考试、斯坦福大学医学院临床推理考试,等等。ChatGPT 的学习、转化、创意能力可以替代编程、翻译、广告策划、财经分析、文学创作、学术写作等 19% 的工作岗位及其吸纳的 80% 的劳动力<sup>③</sup>。同时,ChatGPT 在海量抓取个体数字足迹训练模型的过程中,不仅累计耗费 1287 兆瓦时电能、排放 552 吨二氧化碳当量(近乎每月平添 17.5 万人的耗能量)<sup>④</sup>,而且极有可能被犯罪分子滥用、盗用,进而滋长各类恶意软件、勒索软件、虚假新闻、知识剽窃手段、网络钓鱼活动、商业电子邮件入侵活动和黑客社会工程学,等等<sup>⑤</sup>。因此,美国《连线》杂志发表评论称,ChatGPT 掩藏的社会威胁和环境成本将构成巨大的“进步陷阱”,有必要强制那些一再罔顾自身责任、不断炮制技术神话的创新者对技术决策负责<sup>⑥</sup>;美国《华盛顿邮报》抨击 OpenAI 一意孤行地沉迷于以快速迭代模型圈占用户的鲁莽竞赛,毫不在乎其自私计划会以牺牲人类安全为代价,而解决

这一问题的办法是负责任创新<sup>⑦</sup>;加拿大滑铁卢大学批判性媒介实验室主任马修·欧高曼(Marcel O’Gorman)指出,山姆·阿尔特曼(Sam Altman)应从肆意改造社会的技术闪电战中悬崖勒马,令 ChatGPT 及其引发的“人工智能军备竞赛”接受伦理和人权约束,重拾负责任创新原则<sup>⑧</sup>。

在人工智能冲击公共利益屡见不鲜的当下,猝然勃兴的 ChatGPT 难以像互联网产业史上的“网景时刻”“iPhone 时刻”般激起一致赞许,相反,越来越多的人开始关切如何诉诸负责任创新原则为 ChatGPT 设置规制。值此之际,大众超乎以往的忧惧倾向绝非凭空产生,因为数据消耗量、能源消耗量均达到空前规模的 ChatGPT 与传统人工智能相比,显著特征是拥有足够实力凭借大语言模型缔造通用人工智能(Artificial General Intelligence, AGI),导致 ChatGPT 可能把传统人工智能在不同应用领域引起的风险聚合,由此催生全方位、多层次渗透数字社会的总体性风险,使深陷滞后效应的监管问责体系更加疲于应对,以至于彻底走向崩溃<sup>⑨</sup>。若继续宽宥各种训练 ChatGPT 的创新活动肆意演化,无异于默许数字社会的总体性风险加速蔓延和变相鼓励互联网巨头轮番开展攀比谁更擅长的竞次(Race to the Bottom)游戏<sup>⑩</sup>。但从人工智能伦理评估人员被互联网公司解聘,到人工智能创新者公然罔顾外界问责,种种迹象表明引导 ChatGPT 实现负责任创新面临巨大阻碍<sup>⑪</sup>,使追问如何遏止

- ① S.Shankland.AI is Speeding Us Toward Intelligent Computers and the Singularity, Pioneer Says, <https://www.cnet.com/tech/computing/ai-is-speeding-us-toward-intelligent-computers-and-the-singularity-pioneer-says/>.
- ② F.Agomuoh.ChatGPT:How to Use the AI Chatbot Everyone’s Talking About, <https://www.digitaltrends.com/computing/how-to-use-openai-chatgpt-text-generation-chatbot/>.
- ③ M.Kan.OpenAI:ChatGPT Could Disrupt 19% of Us Jobs, is Yours on the List?<https://www.pcmag.com/news/openai-chatgpt-could-disrupt-19-of-us-jobs-is-yours-on-the-list>.
- ④ K.Ludvigsen.ChatGPT’s Electricity Consumption, <https://towardsdatascience.com/chatgpts-electricity-consumption-7873483feac4>.
- ⑤ C.Winlo.Is ChatGPT a Disaster for Data Privacy?<https://www.business-reporter.co.uk/digital-transformation/is-chatgpt-a-disaster-for-data-privacy>.
- ⑥ A.Birhane, D.Raji.ChatGPT, Galactica, and the Progress Trap, <https://www.wired.com/story/large-language-models-critique/>.
- ⑦ D.Walker, H.Taneja.AI Changes Everything, We Need New Guardrails to Survive It.And Soon, <https://www.washingtonpost.com/opinions/2023/03/28/ai-machine-learning-safeguards/>.
- ⑧ M.O’Gorman.The AI Arms Race Highlights the Urgent Need for Responsible Innovation, <https://theconversation.com/the-ai-arms-race-highlights-the-urgent-need-for-responsible-innovation-200218>.
- ⑨ N.Helberger, N.Diakopoulos.ChatGPT and the AI Act, Internet Policy Review, 2023 ( 1 ).
- ⑩ C.Kang, A.Satariano.AI Regulation can be Puzzle to Lawmakers, The New York Times, 2023-03-04 ( A1 ).
- ⑪ S.Harrer.Attention is Not All You Need:The Complicated Case of Ethically Using Large Language Models in Healthcare and Medicine, eBioMedicine, 2023 ( 4 ).

ChatGPT 与负责任创新的伦理框架一再脱节成为当务之急<sup>①</sup>。为反思上述问题, 笔者从梳理负责任创新的核心内涵出发, 分析 ChatGPT 迈向负责任创新面临阻碍的来源及具体表征, 并探讨 ChatGPT 转向负责任创新的可能出路。

## 二、作为技术伦理准则的负责任创新

回顾负责任创新的概念史可知, 其可追溯到 20 世纪 70 年代西方学术界围绕如何教育科学家展开的跨学科讨论。彼时, 科学技术对霸权主义、强权政治、环境污染的助推作用不断凸显, 西方学术界认为这与传统科学教育模式重物质轻人文、重专业独立性轻技术溢出效应的偏狭特征密切相关, 因此, 有必要令科学教育转向开放、包容、以人为本的“文脉学习”(Contextual Studies)——把科学技术放归社会结构中权衡利弊, 唤醒科学家为人类命运负责的良心<sup>②</sup>。此后的 20 余年间, “文脉学习”提倡的责任观被科学与技术的人文社会学研究 (Science and Technology Studies, STS) 充分发扬, 后者着重解构现代经济理论把科学技术的创新言说成永远无懈可击、必能促进繁荣与发展的神话, 认为这种神话致力于帮助创新过程赢取免遭质疑、问责和干预的特权, 以便摆脱伦理准则、公众讨论和风险防范措施等外部影响, 因此, 亟须将其纳入有关创新治理活动 (即“创新治理”, Innovation Governance) 加以事先的“社会塑造”<sup>③</sup>。随着科学和技术的人文社会学研究逐渐衍生政策感召力, 欧盟委员会较早接受并积极贯彻让创新者承担社会责任的理念。2011 年, 欧盟委员会科学政策小组组长与开放科学项目负责人勒内·冯·朔姆伯格 (René von Schomberg) 率先把负责任创新较为明确地界

定为: “一个透明的互动过程, 通过这个过程, 其他的社会行动者和创新者相互响应, 以创新过程及其适销产品的伦理可接受性、可持续性和社会的悦纳性 (Societal Desirability) 为目标, 让科学技术的进步成果得以恰当地嵌入社会。”<sup>④</sup>

受欧盟影响, 负责任创新开始作为技术伦理准则得到空前的讨论和普及<sup>⑤</sup>。2013 年, 时任英国埃克塞特大学商学院院长的理查德·欧文 (Richard Owen) 以英国工程与物理科学研究委员会 (Engineering and Physical Sciences Research Council, EPSRC) 为阵地, 联合多学科专家奠定了迄今为止广受国际学术界认可的负责任创新框架。在《负责任创新: 管理社会中负责任地涌现的科学和创新》(Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society) 一书中, 欧文研究团队把负责任创新的概念内涵泛化为“一种通过对当下的科学创新采取响应式管理 (Responsive Stewardship) 从而关怀未来社会的集体承诺”, 并据此提出由四大维度构成的概念模型, 即 ARDR<sup>⑥</sup> (见表 1)。

上述模型提出后在被英国国家科研与创新署 (UK Research and Innovation, UKRI) 采纳为正式制度的同时, 被欧盟委员会认可并向更大范围推广<sup>⑦</sup>。截至 ChatGPT 问世前, 不乏欧洲、北美洲学者与公共机构根据该模型探索核能技术、纳米技术、化工技术、生物医学、食品安全、森林管理等议题, 但上述研究转化、落实的效果并不理想。尽管 ARDR 模型设计的四个步骤在一定程度上可以约束政府主导的自上而下型创新活动, 但当其面临高度商业化、投机化的全球技术市场时暴露了一个重大缺陷——把负责任创新的落实过程简化为按部就班、循序渐进的线性流程, 无法应对众多游离固

① M.J.Kumar.ChatGPT is Not What You Think It is, IETE Technical Review, 2023 (1).

② M.Brown.The Social Responsibility of the Scientist, New York: Free Press, 1971: 1-10.

③ M.Van Oudheusden.Where are the Politics in Responsible Innovation? European Governance, Technology Assessments, and Beyond, Journal of Responsible Innovation, 2014 (1).

④ R.Von Schomberg.Prospects for Technology Assessment in a Framework of Responsible Research and Innovation. M.Dusseldorp, R.Beecroft, et al. Technikfolgen Abschätzen Lehren: Bildungspotenziale Transdisziplinärer Methoden, Wiesbaden: Springer VS, 2012: 39-61.

⑤ NWO.Ethics as a Compass for EU Innovation, <https://www.nwo.nl/en/cases/ethics-compass-eu-innovation>.

⑥ R.Owen, J.Stilgoe, P.Macnaghten, et al.A Framework for Responsible Innovation.R.Owen, J.Bessant, M.Heintz, et al. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society, Chichester: John Wiley & Sons, Ltd, 2013: 36-39.

⑦ R.Owen, M.Pansera, P.Macnaghten, et al.Organisational Institutionalisation of Responsible Innovation, Research Policy, 2021 (50).

表1 负责任创新的ARDR模型表

| 概念维度                   | 具体所指  |
|------------------------|---|
| 预期 (Anticipatory)<br>↓ | 诉诸经验推断、技术评估、情景设计等方法, 描述分析某种创新在经济、社会、环境等方面有望产生的影响和可能的意外, 尽量筛除决策盲点。 |
| 反思 (Reflective)<br>↓   | 反思某种创新潜在的意图、动机和影响是否违拗已知的规章制度、伦理评价, 是否会滋生未知的风险、盲区、悬疑和两难困境。         |
| 慎议 (Deliberative)<br>↓ | 通过对话、参与和辩论, 倾听不同利益相关者对某种创新观点, 进而引入更宽广的视角重新界定重要议题, 并确定潜在涉及争端的领域。   |
| 响应 (Responsive)        | 通过有效的参与性、预期性治理机制, 确定后续创新方向、轨迹和步调, 使创新成为迭代的、包容的、开放的且有动态能力的自适应学习过程。 |

定体制之外的创新活动<sup>①</sup>。因此, 面对发展迅猛的 ChatGPT, 长期呼吁负责任创新的学者如梦初醒、措手不及。事实上, 从 2018 年 GPT-1 诞生到 2023 年 GPT-4 引发国际舆论震惊, ARDR 模型试图建构伦理约束的每一步都被 OpenAI 及其支持者规避。英国伦敦政治经济学院管理学教授莱斯利·威尔科克斯 (Leslie Willcocks) 指出, 在 ChatGPT 一骑绝尘的创新活动背后, 关于负责任的伦理诉求不仅反应迟钝, 而且已被排挤得近乎完全缺席<sup>②</sup>。ARDR 模型究竟忽视了何种关键因素, 以致其落实过程面临巨大阻碍? 需要回归创新活动未被 ARDR 模型审视的基础伦理逻辑加以分析。

### 三、竞争逻辑对 ChatGPT 迈向负责任创新构成的阻碍

在市场决定资源配置的当代社会, 创新活动之所以能够赢得近乎无懈可击的正当性, 一定程度在于其基础伦理逻辑——竞争逻辑被认为是让社会

保持活力的文明机制 (Civil Mechanism)<sup>③</sup>。美国芝加哥经济学派奠基人弗兰克·海尼曼·奈特 (Frank Hyneman Knight) 较早从历史经验中归纳出该取向, 并在 20 世纪 20 年代发表的《竞争伦理学》(The Ethics of Competition) 一文中指出, 竞争作为一种试图赶超他人以夺取某类重要事物的动机, 足以让社会效率越来越高, 被当代社会赋予伦理上的优势地位<sup>④</sup>。创新经济学奠基人约瑟夫·阿洛伊斯·熊彼特 (Joseph Alois Schumpeter) 把竞争视为改善陈腐经济结构的创造性破坏 (Creative Destruction), 后者作为创新的引擎拥有最出类拔萃的卓越品质<sup>⑤</sup>。诸如此类的价值判断在 20 世纪末成为西方主流意识形态, 深刻影响了负责任创新的伦理旨趣。譬如, 强烈支持负责任创新的欧盟认为, 技术竞争能让创新本身成为目标并通过自由市场改善产品和服务, 因此, 负责任创新不应改变竞争逻辑, 应继续发扬竞争逻辑推进社会责任<sup>⑥</sup>。虽然这种主张看似带有“既要竞争又要负责”的辩证色彩, 但实际与此前的经济学观点一样预

① S.Jakobsen, A.Fløysand, J.Overton. Expanding the Field of Responsible Research and Innovation (RRI) —From Responsible Research to Responsible Innovation, European Planning Studies, 2019(12).

② L.Willcocks. Where has Our Ethical and Social Responsibility Gone with ChatGPT? <https://blogs.lse.ac.uk/businessreview/2023/04/03/where-has-our-ethical-and-social-responsibility-gone-with-chatgpt/>.

③ C.Lütge. The Ethics of Competition: How a Competitive Society is Good for All, Northampton, MA: Edward Elgar Publishing, 2019: 2-30.

④ F.Knight. The Ethics of Competition, The Quarterly Journal of Economics, 1922(4).

⑤ Y.Shionoya. The Ethics of Competition, European Journal of Law and Economics, 1995(2).

⑥ R.Schomberg. A Vision of Responsible Research and Innovation. R.Owen, J.Bessant, M.Heintz, et al. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society, Chichester, UK: John Wiley & Sons, 2013: 57-58.

设竞争逻辑完美无缺,忽视了竞争逻辑与社会责任间的矛盾难以在各自维持原状的前提下被直接调和。原因在于市场经济意义上的竞争逻辑的本质是“存在利益冲突的企业各自动用一套有组织的资源和行动,通过提高自身绩效来贬损敌手绩效……永无止息地采取进攻性战略和防守性反击,以确保比敌手牟取更多利益”<sup>①</sup>,与社会责任论反对绩效至上主义、要求企业致力于创造共享价值并为一切利益相关者(不仅仅是股东)服务的主张<sup>②</sup>堪称方枘圆凿。事实上,如果负责任创新伦理不能推动竞争逻辑作出必要调整,那么,人际竞争、公司竞争、国家竞争将同时对“负责任”规制创新的尝试构成巨大阻碍。

### (一) 人际竞争阻碍个体负责

传统创新伦理把个体创新者发动人际竞争称为对弘扬企业家精神(Entrepreneurship)的努力,高度肯定个体创新者拥有开辟自我王国的远大梦想,彰显了耻居人后、力争上游的坚定意志,渴望享受伸张内在精力和聪明才智的纯粹乐趣<sup>③</sup>。该论调长期以“天才崇拜”的形态广泛流行,诱导社会对个体创新者报以宽容、艳羡,进而选择性地忽略其逞才使气、孤标傲世乃至全然不顾一切后果的偏执倾向<sup>④</sup>。在上述局面下,个体创新者难以产生对行为的外部性负责的自觉,通常坚信自身绝对正确,拒斥谨慎反思、斟酌如何在创新过程中平衡个体利益与集体利益、短期利益与长远利益。

2019年,ChatGPT的主要创始人山姆·阿尔特曼撰文宣传“极端的人能够赢取极端的收获”,认

为个体创新者只有驰骋野心、任性妄为、挑战规则、大胆投下有望赚取100倍利润的赌注才能持续战胜竞争对手,最终做到“以惊人的尺度勒令世界屈服于你的意志”<sup>⑤</sup>。但当ChatGPT引发全球范围大规模技术风险时,阿尔特曼一方面展现关切人工智能发展速度过快的姿态,另一方面毫不掩饰自己渴望永远领先竞争对手、牢固主导全球人工智能行业风向的野心<sup>⑥</sup>,反复抛出“用人工智能消灭工作的必要性”等乌托邦预言,为其后续行动辩护,否定来自技术伦理学领域的质疑和批判,并号召他人由衷接受并积极配合其行动计划<sup>⑦</sup>。

阿尔特曼对人际竞争的迷恋极具代表性地印证了企业家精神的固有缺陷:把个体成功视为重中之重,助长自我中心主义情结<sup>⑧</sup>。这种缺陷被倚重竞争逻辑的负责任创新伦理沿袭,使后者关于“既要竞争又要负责”的倡议在个体创新者层面缺乏现实可操作性,通常被追求唯我独尊而非反躬自省、克己自律的人际竞争架空。不仅如此,由于企业家精神的伦理正当性已经根深蒂固,即使阿尔特曼等个体创新者公然以“建设数字乌托邦”或“靠竞争实现自我梦想”为借口推卸责任,也难以被长久受竞争逻辑濡染而惯于单向度肯定、追捧个体奋斗神话的公众及时反制——正如ChatGPT勃兴后,围绕阿尔特曼本人衍生的模因迅速吸引为数众多且与日俱增的狂热拥趸<sup>⑨</sup>。

### (二) 公司竞争阻碍行业负责

负责任创新的ARDR模型呼吁技术行业在创新过程中践行德性原则、透明原则和包容原则,将对社会的关怀与对技术的研发进行整合,共

① A.Minà,G.Dagnino,G.Vagnani.An Interpretive Framework of the Interplay of Competition and Cooperation,Journal of Management and Governance, 2020(24).

② A.Carroll.Corporate Social Responsibility:Perspectives on the CSR Construc's Development and Future,Business & Society, 2021(6).

③ J.Schumpeter.The Theory of Economic Development,Cambridge,MA:Harvard University Press, 1934:93.

④ A.Marwick.Silicon Valley isn't a Meritocracy, and It's Dangerous to Hero-worship Entrepreneurs,https://www.wired.com/2013/11/silicon-valley-isnt-a-meritocracy-and-the-cult-of-the-entrepreneur-holds-people-back/.

⑤ S.Altman.How to be Successful,https://blog.samaltman.com/how-to-be-successful.

⑥ K.Wendt.What can We Learn From Lex Fridman's Interview With Sam Altman?https://www.lesswrong.com/posts/tyts4Dw7SafsxBjar/what-can-we-learn-from-lex-fridman-s-interview-with-sam.

⑦ D.Ingram.Sam Altman is Tech's Next Household Name —If We Survive the Killer Robots,https://www.google.com.hk/amp/s/www.nbcnews.com/news/amp/rcna70371.

⑧ A.Örtenblad.Against Entrepreneurship:A Critical Examination,Cham,Switzerland:Palgrave Macmillan, 2020:62-63.

⑨ J.Funk,G.Smith.Don't Believe the Hype:Why ChatGPT is Not the“Holy Grail”of AI Research,https://www.salon.com/2023/03/19/dont-believe-the-hype-why-chatgpt-is-not-the-holy-grail-of-ai-research/.

同遵循行业自律守则等软性法治工具(Soft Law Instruments)<sup>①</sup>。但上述愿景难以在蹈袭创新经济学的竞争逻辑条件下落地,创新经济学语境中作为创造性破坏的竞争属于行业内的替代性竞争,其伴随掌握不同创新成果的公司之间近乎你衰我兴、你去我留的零和博弈<sup>②</sup>。当业内公司针对特定技术创新陷入“战阵之间,不厌诈伪”的倾轧、算计时,率先开展技术创新活动的公司既试图淘汰竞争对手,又时刻面临竞争对手发起的“商业盗窃效应”(Business Stealing),此时谈论行业的集体负责并不现实<sup>③</sup>。

尽管微软执行副总裁布拉德·史密斯(Brad Smith)针对ChatGPT发布公司通告《迎接人工智能时刻:通过负责任的人工智能推动未来》(Meeting the AI Moment: Advancing the Future through Responsible AI),称“六年来,微软一直致力于投资跨公司的合作项目,以确保我们的人工智能系统可以被负责任地设计”<sup>④</sup>,但实际并非如此。2023年1月,微软向OpenAI投资约100亿美元,将ChatGPT集成到Bing搜索引擎和Edge浏览器中,矛头直指搜索引擎市场,谷歌迅速推出功能近似ChatGPT的聊天机器人Bard。与此同时,You.com、DuckDuckGo和Neeva等初创公司把聊天机器人与本公司的图像搜索、社交应用相结合,纷纷在人工智能竞争中“跑马圈地”,但该行为不仅遭到微软的“报复性威胁”,而且进一步刺激微软更严密地阻挠外界访问其人工智能数据库<sup>⑤</sup>。正如媒介学者大卫·卡普夫(David Karpf)所言,上述迹象反映了互联网行业正在以参加军备竞赛的心态研发ChatGPT及其仿制品——公司真正关心的不是出于对用户利益的保护而加强对可

能有害的后果的约束,而是致力于如何让创新服务自己、碾轧任何试图瓜分行业利润的对手<sup>⑥</sup>。长此以往,整个行业将无法实现对任何公司的有效监管。

事实上,谷歌曾对其提供的人工智能产品在使用范围方面予以限制,一度较为谨慎地考量聊天机器人撰写虚假消息、复制社会偏见并被滥用等问题,但随着微软对谷歌市场份额的步步紧逼,谷歌修改了谨慎的研发战略,尽其所能地避免落后<sup>⑦</sup>。显然,在负责任创新伦理全盘接受竞争逻辑的现状下,行业内强势的创新诉求会表征为同行竞争压力导致有关负责任创新的诉求被束之高阁,即使以创新为名上演的竞争活动打着兼顾社会责任的旗号,最终也难免流于一纸空文。

### (三) 国际竞争阻碍制度负责

对技术创新活动而言,如果人际竞争、公司竞争诱发的个体与行业不负责任被视为市场失灵症候,那么,政府通过宏观调控树立竞争制度将有望强制“创新”与“负责任”重归耦合。但随着技术创新越来越成为国际竞争的重要筹码,部分国家为利用新技术打压他国或出于防范新技术被他国利用等目的,通常或主动、或被迫地对个体与行业的不负责任加以宽纵甚至沦为同谋,从而使有望弥补市场失灵的制度负责局面缺席。

以ChatGPT为例,微软、谷歌呈恶性循环趋势的人工智能军备竞赛充分凸显了政府监管的必要性,美国技术伦理组织“人工智能和数字政策中心”(Center for Artificial Intelligence and Digital Policy, CAIDP)公开抗议ChatGPT存在偏见、欺骗性并严重威胁公民隐私和公共安全,要求政府禁止OpenAI发布GPT-4的最新商用版本,并制定政

- ① R.Lee, J.Petts. Adaptive Governance for Responsible Innovation. R.Owen, J.Bessant, M.Heintz, et al. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society, Chichester, UK: John Wiley & Sons, 2013: 154-155.
- ② C.Ziemnowicz, A.Joseph. Schumpeter and Innovation. G.Carayannis, et al. Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship, New York: Springer, 2013: 1173-1175.
- ③ D.Garcia-Macia, C.Hsieh, P.Klenow. How Destructive is Innovation? Econometrica, 2019(5).
- ④ B.Smith. Meeting the AI Moment: Advancing the Future Through Responsible AI, <https://blogs.microsoft.com/on-the-issues/2023/02/02/responsible-ai-chatgpt-artificial-intelligence/>.
- ⑤ J.Kaufman. Microsoft Wants to Ban Artificial Intelligence, but Not for Your Safety, <https://crast.net/298578/microsoft-wants-to-ban-artificial-intelligence-but-not-for-your-safety/>.
- ⑥ T.H.Tran. The Chatbot Wars Will Change the Internet Forever, <https://www.thedailybeast.com/google-and-microsofts-chatbot-war-might-change-the-internet-forever>.
- ⑦ A.Mudaliar. Google Raises Alarm Over ChatGPT—To Unveil New AI Chatbot in Response, <https://www.spiceworks.com/tech/artificial-intelligence/news/google-red-alert-chatgpt-ai-chatbot/>.

策保护消费者和其他中小公司<sup>①</sup>。但美国部分政府官员秉持极端的冷战思维，极力传播一种谬论：中国科技公司在与美国的人工智能竞赛中具有优势的原因在于未受到监管；而美国政府只要严格监管人工智能，就会向中国让出技术优先地位<sup>②</sup>。此类鹰派人物的谬论被用来反对美国政府对科技巨头的监管，助推了前文所言的竞次游戏。

2023年3月，当生命未来研究所（Future of Life Institute, FLI）组织数千名人工智能研究人员和人工智能伦理学者签署呼吁暂停生成式人工智能发展的公开信时，部分美国人工智能资深人士一方面承认不存在现实的方法实施暂停，另一方面指出如果美国按下暂停键，将使中国获得优势<sup>③</sup>。2023年4月11日，GPT-4推出不到1个月，中国国家互联网信息办公室发布《生成式人工智能服务管理办法征求意见稿》，针对ChatGPT和类似技术产品可能造成的危害（侵犯国家安全、泄露公民隐私、捏造虚假信息、煽动不正当商业营销和竞争等）提出治理措施，要求创新者打开算法“黑箱”，帮助用户科学认识产品的技术原理，防止沉迷其中，同时，强制任何破坏法治与公序良俗的创新者暂停或终止服务<sup>④</sup>。这与欧盟通过《人工智能法案》采取的规制不同，与美国的方法也不同，现阶段，后者尚缺乏全面的立法或监管议程。

反观曾为负责任创新先行者的英国，“脱欧”后未能坚持往日主张，在急切奔赴国际竞争的道路上愈发将“创新”与“负责任”割裂。2023年3月29日，英国的科学、创新和技术部（Department for Science, Innovation and Technology, DSIT）为应对ChatGPT造成的冲击颁布《偏倚创新的人工智能规制途径》（A Pro-Innovation Approach to AI Regulation），明确要把英国变成引领全球人工智能竞争的超级大国。因此，其一方面

施行宽松的监管制度，另一方面斥资200万英镑用沙盒实验法评估已出台和拟出台的制度对创新活动可能构成的阻碍，最大限度地剔除这些阻碍<sup>⑤</sup>。上述举措貌似在瞄准国际竞争的同时促成一定程度的制度负责，但实际是将制度本身等同“应删尽删”的赘余物——不仅使如何将人工智能应用于招聘、教育和就业等尚未得到全面监管的领域显得愈发模糊，而且可能导致英国监管机构虽然面临空前繁重的人工智能问题，却吊诡地承担较以往任何时代都要少的社会责任<sup>⑥</sup>。

从世界重要人工智能大国的反应中可以发现：围绕ChatGPT展开的国际竞争俨然构成势头强劲的干扰性外力，其强迫各国政府把治理本土技术市场失灵的內政按照外务的规格进行处理。此时，理应自主、有序、谨慎、稳健推进的负责任创新计划将迅速让位于利益压倒伦理的国际竞争，导致矫正个体与行业不负责任的制度延迟出台。

#### 四、转向共生逻辑：ChatGPT实现负责任创新的可能出路

综合前文以ChatGPT为典型例证的分析不难发现：作为技术伦理准则的负责任创新，由于完整继承了创新经济学的竞争逻辑，导致“创新”与“负责任”的固有矛盾未得到有效调和——使ARDR模型构想的线性负责任流程极易被高速驱策技术迭代的人际竞争、公司竞争、国际竞争扰乱，反复造成“负责任”向“创新”的无度妥协。因此，要推动负责任创新伦理对ChatGPT产生真正有效的约束力，亟须修正强势奠定负责任创新基调的竞争逻辑。如果选择竞争以外的技术发展道路，是否注定无法保持创新？换言之，负责任创新中的“创新”，是否只能是熊彼特所谓的靠竞争驱动的

① CAIDP. In the Matter of OPEN AI (FTC 2023), <https://www.caidp.org/cases/openai/>.

② P.Triolo. ChatGPT and China: How to Think About Large Language Models and the Generative AI Race, <https://thechinaproject.com/2023/04/12/chatgpt-and-china-how-to-think-about-large-language-models-and-the-generative-ai-race/>.

③ R.Garver. Call for Pause in AI Development may Fall on Deaf Ears, <https://www.voanews.com/a/call-for-pause-in-ai-development-may-fall-on-deaf-ears/7029969.html>.

④ 《国家互联网信息办公室关于〈生成式人工智能服务管理办法（征求意见稿）〉公开征求意见的通知》，[http://www.cac.gov.cn/2023-04/11/c\\_1682854275475410.htm](http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm).

⑤ DSIT. A Pro-Innovation Approach to AI Regulation, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1146542/a\\_pro-innovation\\_approach\\_to\\_ai\\_regulation.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146542/a_pro-innovation_approach_to_ai_regulation.pdf), March 2023.

⑥ R.Morrison. New UK AI Regulation White Paper Leaves “Unanswered Questions” on ChatGPT, <https://techmonitor.ai/technology/ai-and-automation/uk-ai-regulation-chatgpt>.

“创新”？

该问题在西方主流经济学者遵奉的熊彼特主义创新范式内无从索解，中国学者给出另一种可能，即转向共同富裕愿景下的后熊彼特主义创新范式，据此重振负责任创新伦理的有效性。上述创新范式的内涵可概括为：“从企业家个体驱动的创新转向用户创新、创客创新为主。企业创新的主导使命从经济使命主导转向政治、经济、社会、科技等多使命主导；企业间的竞争逻辑从竞合逻辑主导转向价值共创共生逻辑主导；以数字平台主导的平台创新生态系统逐步从平台‘赢者通吃’逻辑转向‘共享赋能’逻辑……从市场逻辑驱动的创业企业家精神转向共益型企业家精神；组织创新使命层面需要从管理控制、效率至上转向赋权个体、知识创造转化。”<sup>①</sup>后熊彼特主义创新范式既反对把公司盈利视作创新的主导目标，又反对把社会责任变成公司间靠“行善赚钱”助推市场竞争的工具，强调把社会责任化作创新的内驱力，使创新过程立足社会价值捕获、社会价值创造、社会价值分配，“以社区、社群乃至社会生态圈为基础的共益型企业家精神驱动企业实现新一轮的共赢增长”<sup>②</sup>。与熊彼特主义创新范式相比，后熊彼特主义创新范式的独特性见表2。

总体而言，以共生逻辑置换竞争逻辑构成后熊

彼特主义创新范式的内在规定性。这种共生逻辑指公司之间、公司和诸多利益相关者之间在创新过程中开放边界，超越传统竞争关系的零和博弈，从非此即彼走向彼此需求、彼此依存、互动协作、和谐共存、协同进化，创造一个实现协同价值的共同繁荣的生态群落<sup>③</sup>。只有把立足共生逻辑的“创新”与“负责任”重新连接，才能真正抵消双方固有的矛盾，进而合成对人际竞争、公司竞争、国际竞争发挥有效抗衡作用的负责任创新伦理。具体到ChatGPT议题，转向共生逻辑的负责任创新伦理可以通过两类途径规制日趋军备竞赛化的无序竞争。

#### （一）削弱不负责任创新赖以扩散的社会基础

如前文所述，ChatGPT之所以在面临众多经济风险、政治风险、文化风险和環境风险的情况下依然能在2个月内俘获1亿用户，成为史上用户增速最快的消费级应用程序，一个重要原因是竞争逻辑已被建构成保障现代文明永续发展的神话，使人际竞争、公司竞争驱动的不负责任创新极易得到公众和某些国家的宽纵与追捧。对此，有研究指出，当今社会习惯把各种野蛮生长的器物创新错认为“进步”，进而将ChatGPT等价于不容矫治的必然趋势，充分暴露了丧失批判性思维的技术创新拥趸的广泛存在，后者以拥抱“福音”

表2 熊彼特主义创新与后熊彼特主义创新区别表

| 创新要素 | 熊彼特主义创新          | 后熊彼特主义创新         |
|------|------------------|------------------|
| 创新主体 | 个体精英             | 群体大众             |
| 组织载体 | 纯商业组织            | 混合型组织            |
| 主导逻辑 | 竞争逻辑             | 共生逻辑             |
| 制度支撑 | 重点扶持有某一特征的企业政策   | 使各类企业共享创新资源的普惠政策 |
| 价值导向 | 创业型企业家精神：股东利润最大化 | 共益型企业家精神：为全民发展负责 |
| 创新战略 | 市场本位的经济型战略       | 综合价值与共享价值本位的战略   |
| 创新过程 | 最大程度配置市场经济资源     | 增益整体的经济与社会资源的价值  |
| 创新目标 | 赢者通吃，最大化自身经济利益   | 建成创新成果共享共益的社会生态圈 |

① 陈劲：《探索共同富裕战略的企业创新范式》，《清华管理评论》，2021年第10期。

② 阳镇、陈劲：《迈向共同富裕：企业社会责任的底层逻辑与创新方向》，《清华管理评论》，2022年第Z1期。

③ 辛杰、卞江：《共生视阈下平台企业社会责任评价体系构建》，《经济管理》，2022年第6期。



的热忱张扬着对技术的依赖情结：认定技术加速进步，而法律、伦理使进步变得低效率；认定靠技术创新能做到的事就是应当去做的事；认定技术创新和创新者本身是好的，若存在危害也只是使用者造成的意外<sup>①②</sup>。当前，只有使数以亿计的技术创新拥趸首先在精神世界摆脱对 ChatGPT 及其设计师的盲目接纳、一味追捧，才有望使其在现实世界中从集体助推 ChatGPT 野蛮生长的行动者网络中抽身——在这一点上，转向共生逻辑的负责任创新伦理大有可为。

传播学研究揭示了技术创新成果的扩散绝非“自然”或“必然”的过程，即使这些成果具有便于被潜在用户观察、实操的相对优势，也只有在和意识产生价值兼容性（Value Compatibility）的前提下才能顺利掌握其赖以存续、增殖、繁荣的社会基础<sup>③</sup>。截至目前，ChatGPT 俘获了技术创新的五类采用者（创新者、早期采用者、早期大众、晚期大众与落后者）中占比 2.5% 的创新者，其在未来 6-12 个月内能否迈过早期市场和主流大众市场之间的鸿沟，赢得实用主义者的支持至关重要，这决定了其能否真正成为主导主流市场的“杀手级”应用<sup>④</sup>。如果基于竞争逻辑的负责任创新伦理未能有效引导早期创新者谨慎对待 ChatGPT，那么，有必要在 ChatGPT 跃居“杀手级”应用前及时发起“启蒙 2.0 运动”普及转向共生逻辑的负责任创新伦理，使负责人和广大群众意识到亟须这种新伦理“去挑战市场、官僚机构和科学技术发展的主导逻辑。仅仅因为某些东西可以出售并不意味着它应该被出售；仅仅因为某些东西会被开发并不意味着它应该被开发”，否则“我们会因为过多地将信念投入科技领域而丧失人性或对人权的信念，因此被蒙蔽双眼”<sup>⑤</sup>。经由这种伦理认知，有望改变对 ChatGPT 及其创新者的单向度肯定与纵容倾向——该局面将削弱不负责任创新赖以扩散的社会基础。

（二）使通过负责任停滞促成制度跟进变为可能仅凭“启蒙 2.0 运动”普及转向共生逻辑的负责任创新伦理并不充分。首先，任何思想运动都需要时间才能深入人心，ChatGPT 一骑绝尘的迭代速率显然拒绝“浪费”时间。其次，伦理准则作为公民间靠自觉、自律约定俗成的秩序关系，若迟迟不能由内而外地渗入行业守则、法律法规等获得集体性权力保障的制度，注定松散脆弱、不堪一击，无法成为集价值引导和规范约束于一体的、具有普遍性和强制性的力量。因此，无论开展思想启蒙，还是覆盖行业、国家和世界体系的制度设计，都亟待令技术向社会让渡时间——这需要转向共生逻辑的负责任创新伦理把负责任停滞（Responsible Stagnation）原则纳入其中，据此为“责任”创造出足以对人产生有效约束力的外部环境。

负责任停滞不是复古地、一刀切地永远禁止或压抑所有技术研发活动，而是指当某些技术研发活动虽会刺激技术种类与经济产值的数量增长，但其利益相关者认为已然 / 或然酿成重大的负外部性时，应暂时搁置原定研发计划，直到技术演化水平与政策监管水平间的滞差得到妥善消减，或直接放弃原定研发计划，让各类利益相关者讨论得出更有可能实现安全的高质量发展（而非危险的高数量增长）的新计划。当转向共生逻辑的负责任创新伦理把负责任停滞纳入内部并作为自身存在的前提时，熊彼特范式为“创新”与“停滞”建构的二元对立关系便被解构。此时，“创新”不再单一地表征为技术种类与经济产值的数量增长；相反，在特定时段内，其能够且必须呈现承担责任而放缓、放弃数量增长的状态<sup>⑥</sup>。从表面看，负责任停滞关于集体反思、商谈的诉求和传统负责任创新伦理的 ARDR 模型相仿，但事实上，前者为“责任”赋予了空前强势的伦理地位（区别于 ARDR 模型只用“责任”修饰熊彼特范式的“创新”，而实际在暗中偏袒后

① E.Dans.ChatGPT and the Decline of Critical Thinking, <https://www.ie.edu/insights/articles/chatgpt-and-the-decline-of-critical-thinking>.

② R.Doshi, S.Bajaj, H.Krumholz.ChatGPT Ethics: Temptations of Progress, <https://bioethicstoday.org/blog/chatgpt-ethics-temptations-of-progress/>.

③ D.Bunker, K.Kautz, A.Nguyen.The Role of Value Compatibility in Information Technology Adoption.B.Donnellan, T.Larsen, L.Levine, et al. The Transfer and Diffusion of Information Technology for Organizational Resilience, Boston: Springer, 2006: 53-70.

④ 方兴东：《ChatGPT 爆红确立智能传播的主流地位》，《中国社会科学报》，2023 年 3 月 7 日。

⑤ 胡泳：《呼唤启蒙 2.0——人类意识的范式转变与革命》，《探索与争鸣》，2020 年第 6 期。

⑥ D.Guston.Responsible Innovation: Who Could be Against That?Journal of Responsible Innovation, 2015(1): 1-4.

者),进而使“责任”不再等待被各种技术研究计划兼顾,而是主动抗衡甚至一票否决曾被认为不推行就不能让社会“进步”的计划。

目前,上述愿景已在呼吁暂停训练 ChatGPT 的运动中初露端倪,但其面临的挑战依然严峻。2023年3月22日,特斯拉首席执行官埃隆·马斯克(Elon Musk)、苹果公司联合创始人史蒂夫·沃兹尼亚克(Steve Wozniak)、人道科技中心首席执行官特里斯坦·哈里斯(Tristan Harris)等1000余名技术精英在致力于防范人工智能风险的非营利机构“未来生命研究所”官网发布联名信,声称影响人类社会命运的技术决策不能委托给未经选举的技术领导者,只有当其确信 ChatGPT 的影响是积极的且风险可控时,才能允许相关研发活动继续。因此,ChatGPT 应被暂停训练至少6个月以便监管措施跟进,如果全球范围内有任何机构或个人一意孤行,所有国家的政府需要强行干预<sup>①</sup>。值得注意的是,2023年4月,马斯克在美国内华达州成立 X.AI 公司,既挖走 DeepMind 旗下的顶级人工智能科学家伊戈尔·巴布什金(Igor Babuschkin),又购置数千个图形处理单元(GPU)为即将推出的生成式人工智能产品 TruthGPT 提供动力<sup>②</sup>。马斯克的这一举动充分证实,负责任停滞局面几乎不可能通过精英集团的口头承诺实现。因此,要使负责任停滞真正融入基于共生逻辑的负责任创新伦理并拱卫后者,应专门制定国内法和国际公约,对执行时间、如何执行等问题作出明确规定。从 ChatGPT 的案例看,美国和英国的立法机构之所以对国内人工智能公司与企业家极尽宽纵,一定程度是因为两国试图通过人工智能军备竞赛推行技术霸权主义——这种图谋促使其他国家试图监管 ChatGPT 的立法者陷入囚徒困境心理,可能迫于压力效仿美国和英国。

制定跨国人工智能治理框架或许应被视作当务之急。如果没有强大且一致的国内、区域和国际治

理安排下的真诚的国际协调,就无法有效解决与人工智能有关的挑战。国际人工智能治理框架有助于营造一致的、对全球有利的监管环境,对于成功地和负责任地采用人工智能技术极为必要。2022年11月18日,欧盟理事会决定支持实施“为和平与安全促进人工智能的负责任创新”项目。欧洲委员会于2018年提出支持人工智能相关研究的指导原则将是发展“负责任的人工智能”,并进一步指出,由于人工智能易于跨国界交易,只有全球解决方案才能在该领域持续发挥作用。欧盟将促进人工智能和一般技术的使用,以帮助解决全球挑战。欧盟表示,其希望为发展“负责任的人工智能”、集体安全以及从人工智能为实现联合国可持续发展目标带来的机遇中获益的潜力作出贡献,并应对人工智能对和平与安全构成的挑战<sup>③</sup>。

2019年,时任联合国秘书长古特雷斯在“AI for Good”全球峰会上指出,如果要利用 AI 的益处并解决风险,就必须一起努力——包括政府、行业、学术界和民间社会——制定框架和制度,以实现负责任的创新<sup>④</sup>。为防范与解决不负责任的行为者转移和滥用人工智能从而带来的和平与安全风险,迫切需要以联合国为中心的国际组织与全球范围内反对技术霸权主义的国家展开协作,共同应对因人工智能创新带来的潜在滥用而引发的人道主义、战略和政治挑战。

## 五、结论与讨论

有学者以“效率优先,兼顾公平”的伦理准则为例,指出当公平的实际发展程度远低于效率时,把双方直接拼贴貌似追求一视同仁,实际是让公平本应得到更多保护才能被克服的先天弱势延续下去<sup>⑤</sup>。与之相似,当熊彼特范式下基于竞争逻辑的“创新”一如既往地占据技术市场中心时,“负责任”即使与其达成名义上的并列,也只是不堪一

① Pause Giant AI Experiments: An Open Letter, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

② M.Marcelline.Elton Musk Creates New Artificial Intelligence Company “X.AI”, <https://www.pcmag.com/news/elon-musk-has-created-new-artificial-intelligence-company-xai>.

③ Council of the European Union.Council Decision (CFSP)2022/2269 of 18 November 2022 on Union Support for the Implementation of a Project “Promoting Responsible Innovation in Artificial Intelligence for Peace and Security”, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022D2269>.

④ United Nations.Secretary-General’s Message for Third Artificial Intelligence for Good Summit, <https://www.un.org/sg/en/content/sg/statement/2019-05-28/secretary-generals-message-for-third-artificial-intelligence-for-good-summit>.

⑤ 吴忠民:《“效率优先,兼顾公平”究竟错在哪里》,《北京工业大学学报(社会科学版)》,2007年第1期。

击、徒有其表的点缀。因此，在以 ChatGPT 挑起的人际竞争、公司竞争、国际竞争中，负责任创新伦理的倡导者始终处于既想约束人工智能军备竞赛又忌惮挫伤增长效率的两难窘境，变相默许“负责任”一次次对“创新”作出权宜性迁就，反复宽纵人工智能军备竞赛的野蛮生长，使负责任创新沦为“约束君子不约束小人”的疲软口号。

作为一项试图用理论影响实践的探索性研究，本文的理论贡献主要包括两点：一是指出在多方势力呼吁（或自称）要使 ChatGPT 合乎负责任创新的意见气候形成后，负责任创新伦理仍无法规制 ChatGPT 的主要原因是 ARDR 模型看似成熟缜密，实则保守地沿袭了熊彼特主义创新范式下基于竞争逻辑的“创新”概念，再将后者与“负责任”概念接合——导致竞争逻辑始终驱动“创新”、凌驾“负责任”之上，使负责任创新伦理存在一定程度的脆弱性、虚伪性。二是指出要实现负责任创新伦理对 ChatGPT 的规制，熊彼特主义创新范式下基于竞争逻辑的“创新”概念应被大幅改造乃至扬弃——具体解决方法是用后熊彼特主义创新范式下基于共生逻辑的“创新”概念另行接合“负责任”，消弭“创新”与“负责任”之间的权势差距，进而建构真正具有普遍约束力的负责任创新伦理准则，据此指导各国出台共同监管 ChatGPT 的法律法规。但笔者深知：转向共生逻辑的负责任创新伦理带来的启发性仍停留在概念范畴，其应结合地方性知识，进而衍生更加细致、务实、可操作的软/硬约束机制——这是后续对策性研究的重中之重，如同大卫·哈维（David Harvey）的呼吁：“话语反思结束的地方，便是实践政治开始的地方。”<sup>①</sup>

笔者无意从事罗列操作细则的对策研究，但希望结合中外时局，提出可供未来讨论的设想与倡议。目前，在 ChatGPT 的全球监管体系迟迟未能建立的宏观背景下，中国推出《生成式人工智能服务管理办法（征求意见稿）》（以下简称《办法》），开全世界监管生成式人工智能之先河，显示出中国对于负责任创新的重视。《办法》第三条指出，“国家支持人工智能算法、框架等基础技术的自主创新、推广应用、国际合作，鼓励优先采用安全可信的软件、工具、计算和数据资源”。未来的对策性

研究应重点探讨中国如何把优化后的生成式人工智能服务管理办法升级为建设人类命运共同体的关键策略，联合深受技术霸权威胁的南方国家及倾向严格监管 ChatGPT 的欧盟，共同推动基于共生逻辑的负责任创新伦理走向世界。

人工智能将从根本上改变人类社会。因为这一转变过程或许对任何国家而言都是不可避免的，所以各国希望能在某种程度上达成合作。人工智能的进展虽可以在诸多层面造福人类，但也成为国家间和国家内部产生或加剧紧张局势、冲突和不安全的新方式。人工智能的某些应用，例如，致命的自主武器系统带来的风险已成为军控界关注的主要问题之一。此外，跟上技术革新的步伐需要大量的技术专长和能力，可能会超出较弱国家的能力并引起严重的权力不对称。与此同时，值得注意的是，人工智能技术可能侵犯人最基本的权利并引发深刻的伦理、社会和法律挑战，因此，当前出现关于各种不确定问题的激烈辩论，这些问题可能会以快速和目前无法想象的方式发生变化，根据经验可知，在此条件下不利于达成国际共识。

目前，竞争驱动的创新促发了对人工智能的错误评估，使人类相信自己可以免受人工智能潜在的负面影响。人工智能军备竞赛为监管者提供了错误的行为激励，因此，在实现监管目标方面的效果较差。负责任的创新作为一种自我管理机制，可以为全球人工智能界提供实用的工具和方法，以确定并有助于预防和减轻人工智能研究和创新的转移、滥用可能对和平与安全造成的风险。促进负责任的创新过程、方法和工具不断完善，是推动人工智能从业人员为减轻上述风险作出贡献的重要途径。鉴于大多数与人工智能相关的政策核心问题是技术，加强政策制定者与技术界的联系尤为重要。同时，有必要提高其他社区对技术创新的认识，教育和动员其更好地参与监管。如果能在全社会形成负责任创新的正确导向，关于人工智能的社会对话水平将得到提升，社会对监管机构设计和维持适当的人工智能监管制度的能力也会强化信任，从而有利于促进人工智能规范的接受度和人工智能实践的合法性。

<sup>①</sup> D.Harvey, Justice.Nature and the Geography of Difference, Cambridge, MA: Wiley, 1996: 438.

---

## Looking at Responsible Innovation From ChatGPT: Impediments and Directions for Transformation

HU Yong ZHU Zheng-de

( School of Journalism and Communication, Peking University, Beijing 100871 )

**Abstract:** The rise of ChatGPT has rekindled the tension between the acceleration of technological innovation and the lagging-behind social adaptation, forcing humanity to face unprecedented economic, political, cultural, and environmental risks and intensifying calls for responsible innovation. However, responsible innovation as a code of technological ethics follows the Schumpeterian paradigm of innovation based on the logic of competition, and its design of the ARDR model, on the one hand, never fundamentally regulates the dominant advantage of innovation over responsibility, and on the other hand, reduces responsibility to a linear process, resulting in the seeming juxtaposition of responsibility with innovation, while in fact it is repeatedly reduced to a mere embellishment in the face of the interpersonal, corporate, and international competition that is embedded in innovation. Only by relying on the symbiotic logic of innovation in defense of responsibility under the post-Schumpeterian paradigm can we construct a truly binding ethical code that will weaken the social basis for the diffusion of irresponsible innovation and make it possible for responsible stagnation to lead to institutional follow-up. Realizing the above goals not only requires countries to improve their internal governance, but also requires responsible powers to develop transnational AI governance frameworks so that the development of AI will be beneficial to all humankind.

**Key words:** ChatGPT ; Responsible Innovation ; Ethical Construction of Technology ; Competitive Logic ; Symbiotic Logic

[ 责任编辑: 王文秋 ]

[ 责任校对: 曹晶晶 ]