doi:10.13582/j.cnki.1672-7835.2023.06.017

ChatGPT 类生成式人工智能监管的国际比较与借鉴

和军,杨慧
(辽宁大学 经济学院,辽宁 沈阳 110036)

摘 要:ChatGPT 类生成式人工智能的快速发展虽在不同领域带来积极影响,但也引发了法律政策、安全技术、社会伦理等层面的监管问题。比较美国、欧盟、日本等国外典型经济体人工智能监管的治理实践,为完善我国人工智能监管治理体系提供政策参考。我国人工智能监管应坚持发展与安全并重、促进创新与依法治理相结合的监管原则,建立健全法律与政策、技术与安全、社会与伦理等层面的监管路径,同时对人工智能进行进入性监管、技术性全程监管和分级分类监管,完善我国的人工智能监管体系。

关键词:ChatGPT;生成式人工智能;政府监管;国际实践;中国路径
中图分类号:D63 **文献标志码**:A **文章编号**:1672-7835(2023)06-0119-10

人工智能是当今科技领域的热点话题,也是未来社会发展的重要驱动力。2022 年 11 月,美国 OpenAI 公司推出 ChatGPT(即 Chat Generative Pre-trained Transformer)人工智能聊天程序,两个月内该聊天程序的活跃用户突破一亿,成为用户增长最快的应用程序,在世界范围内受到广泛关注。ChatGPT 是人工智能技术驱动的自然语言处理工具,具有更强大的人机互动功能,其借助“Transformer 神经网络架构”技术,通过大规模语言模型训练,能够使用类人语言实现对话交流,完成撰写文案、编写代码、编辑脚本、翻译等内容与任务①。ChatGPT 具有非常广阔的应用场景与发展前景,继 ChatGPT 后,百度、阿里巴巴、华为和科大讯飞等互联网巨头纷纷布局 AIGC(artificial intelligence generated content),人工智能时代已然到来。随着人工智能技术的不断创新与应用,ChatGPT 类生成式人工智能对政府监管提出了系列挑战,是我国和国际社会亟需关注的重要问题。

一 ChatGPT 类生成式人工智能的发展与面临的监管风险

(一)ChatGPT 的基本原理

ChatGPT 作为 AIGC 技术的关键性突破,是人工智能技术发展的重要里程碑。国内外研究者对 ChatGPT 基本原理进行了深入探讨。ChatGPT 的工作原理是基于一种机器学习模型(生成式预训练变换器)来生成文本的人工智能应用程序,这种模型在不同阶段通过接收大量的语料库训练,学习并理解人类的语言规则。在训练阶段,模型通过解析数百亿个单词,对人类语言中的各种模式和结构进行建模,进而掌握语言的语法、语义和语境等复杂规则;在生成阶段,通过其在训练期间积累的数据以及对语言的理解进行对输入语言的信息分析,进而生成相应的语言输出②。ChatGPT 的关键技术是通过大规模的语料库进行自我学习的训练方法以实现人类与机器自然交流

收稿日期:2023-08-09
基金项目:辽宁省“兴辽英才”计划项目(XLYC2004012)
作者简介:和军(1972—),男,山西大同人,博士,教授,博士生导师,主要从事政府监管、区域经济研究。

①《GPT-4》,ChatGPT 官网,https://openai.com/research/gpt-4.

②Hanleem A, Javaid M, Singh R P. “An Era of ChatGPT as a Significant Futuristic Support tool: A Study on Features, Abilities, and Challenges”, *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2022, 2(4): 100089.

对话,从而生成更为真实的文本对话内容①。ChatGPT 不断尝试添加图像生成、语音合成、多语言支持②等新的功能模块,通过使用不同指标与评估基准,相较以往人工智能技术更具知识性、情感性、流畅性和多样性③,同时具有拟人化、强交互性和全感官等特征④。

(二)ChatGPT 类生成式人工智能对不同领域带来的发展机遇

ChatGPT 类生成式人工智能技术具有广阔应用前景,为各行业各领域带来新的发展机遇。学者们对 ChatGPT 在行政、金融、零售、教育、医疗保健、卫生⑤、环境⑥等不同领域的发展与应用进行了探讨。总体上,ChatGPT 类生成式人工智能为企业与用户间实时交互提供了全新沟通方式,在帮助企业提高效率和质量的同时降低人力成本,不仅提高了用户满意度,还增加了使用者的忠诚度,为个性化教育、自动化翻译以及智能助手与服务等领域带来机遇⑦。具体而言,在行政领域,人工智能提高了行政治理与公共决策制定的效率,增强了社会治理与风险防控的能力⑧。在金融领域,银行与金融机构通过使用 ChatGPT 虚拟客服服务,更快速有效地回答客户问题,提高服务质量⑨。在零售领域,ChatGPT 可以实现个性化服务、商品营销推荐与销售预测⑩。在教育领域,ChatGPT 提高了教育智能化水平,更好地与学生进行互动答疑,拓展了学习资源与方法,为智能家教与在线教育提供强大技术支撑⑪。在医疗保健领域,ChatGPT 不仅可以帮助医生与患者实时对话,更有效了解病情并提出治疗方案,同时患者也可以借助其完成自我诊断,了解健康生活的医学建议⑫。

(三)ChatGPT 类生成式人工智能引发的监管风险

ChatGPT 类生成式人工智能是一把“双刃剑”,在助力人工智能及相关应用领域发展与产业创新的同时,也为行业监管带来系列风险挑战。

首先,给法律监管带来风险挑战。一是数据泄露与不合规引发的法律风险。ChatGPT 通过对用户数据信息的收集与分析实现内容生成,用户在使用过程中存在无意识情况下输入个人隐私信息、商业机密以及其他敏感数据而引发法律风险的可能,既涉及 ChatGPT 对用户自身的侵权行为,也涉及将这些数据信息用作训练数据而产生的不合规风险。典型的案例为知识产权侵权问题⑬,如版权、著作权、专利权等侵权案件⑭。二是虚假信息带来的法律风险。ChatGPT 存在使用虚假数据与信息,形成虚假观点的情况,这可能会排挤正确观点,对社会安定造成挑战⑮。三是使用者主观故意引发的刑事风险。生成式人工智能技

①Shen Y, Heacock L, Elias J, et al. “ChatGPT and Other Large Language Models Are Double-edged Swords”, *Radiological*, 2023(2): e230163.

②Wu C F, Yin S M, Qi W Z, et al. “Visual ChatGPT: Talking, Drawing and Edit with Visual Foundation Models”, *arXiv preprint arXiv*: 2303.04671, 2023.

③Bang Y, Cahyawijaya S, Lee N, et al. “A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity”, *arXiv preprint arXiv*: 2302.04023, 2023.

④于水,范德志:《新一代人工智能(ChatGPT)的主要特征、社会风险及其治理路径》,《大连理工大学学报(社会科学版)》2023 年第 5 期。

⑤Som S.Biswas. “Role of Chat GPT in Public Health”, *Annals of Biomedical Engineering volume*, 2023, 51(3): 868-869.

⑥Zhu J J, Jiang J Y, Yang M Q, et al. “ChatGPT and Environmental Research”, *Environmental Science & Technology*, 2023(3).

⑦Jianyang D, Yijia L. “The Benefits and Challenges of ChatGPT: An Overview”, *Frontiers in Computing and Intelligent Systems*, 2022, 2(2): 81-83.

⑧刘佳明:《人工智能在行政治理领域应用的挑战及对策》,《领导科学》2023 年第 5 期。

⑨Dowling M, Lucey B. “ChatGPT for Finance Research: The Bananarama Conjecture”, *Finance Research Letters*, 2023, 53(5): 103662.

⑩Justin P, Akiko U, Charles D. “ChatGPT and Consumers: Benefits, Pitfalls and Future Research Agenda”, *International Journal of Consumer Studies*. 2023, 47(4): 1213-1225.

⑪Farrokhnia M, Banihashem S K, Noroozi O, et al. “A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research”, *Innovations in Education and Teaching Intenational*, 2023(3): 1-15.

⑫Sallam M. “ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns”, *Healthcare*, 2023, 11(6): 887.

⑬沈锡宾,王立磊,刘红霞:《人工智能生成内容时代学术期刊出版的机遇与挑战》,《数字出版研究》2023 年第 2 期。

⑭史惠斌,郭泽德:《迈向智能:AIGC 内容生成模式引发的出版变革》,《数字出版研究》2023 年第 2 期。

⑮Mijwil M, Aljanabi M. “Towards Artificial Intelligence-Based Cybersecurity: The Practices and ChatGPT Generated Ways to Combat Cybercrime”, *Iraqi Journal For Computer Science and Mathematics*, 2023, 4(1): 65-70.

术可能被人为恶意利用，通过不当获取数据信息，故意生成违法犯罪信息等利用人工智能技术从事违法犯罪活动等而引发的刑事风险①。

其次，对安全与技术带来风险挑战。ChatGPT 类生成式人工智能以海量数据输入为基础，庞大用户使用量为其提供了大量信息来源，包括用户个人隐私、企业商业数据、甚至国家信息机密，这就存在着潜在安全风险。一是政治安全风险。ChatGPT 类人工智能在使用时的数据信息流转没有国界限制，导致海量用户的数据流入研发企业及其所属国家，在一定意义上数据信息失去自主保护，对国家政治安全造成潜在安全风险。二是算法技术风险。ChatGPT 类人工智能技术的运行核心是算法，而算法模型采用“黑箱”模式，导致人工智能技术具有一定的不可解释性与模糊性。算法在其开发、投入、运行与产出环节的缺陷均可产生潜在风险。如在行政治理领域，存在对个人权利以及由“算法影子”现象产生的对公共秩序与个人权益造成侵害的风险问题②。三是数据与网络安全风险。人工智能核心技术缺陷会引发网络安全方面的潜在风险，如数据信息安全、偏差与欺骗、输出错误信息等③。用户在使用 ChatGPT 过程中输入个人隐私，存在数据泄露风险；亦存在 ChatGPT 编造不实问题而导致决策失误，造成无法估量后果的潜在风险；此外，还存在 ChatGPT 编写代码功能被黑客利用的技术风险，对国家网络安全提出新挑战④。

最后，对经济发展与社会伦理问题带来风险挑战⑤。一是加剧垄断。人工智能技术的应用对数字平台企业具有重要影响，平台企业可能利用算法歧视区别对待合作商家以维护自身的垄断优势。二是劳动力替代。生成式人工智能可以解决部分产业中劳动者的劳动局限问题，提升生产效率的同时也影响着就业岗位的稳定性，进而产生“技术性”失业等劳动力替代问题。三是社会分化加深。人工智能技术与社会生活的紧密融合，一定程度会加剧社会不平等特别是收入不平衡的风险，主要体现在技术创新所带来的社会福利对群体中个体的影响差异。人工智能技术的广泛应用不利于对信息数据、互联网和数字技术掌握与应用能力较弱的群体，加剧“数字鸿沟”，进而加深社会分化⑥。此外，生成式人工智能的类人特征会引发科技伦理与道德层面的其他潜在风险，如误导欺骗、侵犯隐私、价值观渗透和道德缺失等问题⑦。生成式人工智能对其未训练的问题可能提供虚假信息，导致误导欺骗等潜在风险；另外生成内容的可靠、可控均存在不确定性，可能引发科技与社会伦理问题⑧。

二　ChatGPT 类生成式人工智能监管的国际比较与启示

（一）国外主要经济体生成式人工智能监管实践

国际人工智能领域的竞争已从技术与产业应用扩展到国际规则的制定，尤其是人工智能的监管与治理规则。由于 ChatGPT 类人工智能强大的技术能力以及可能带来的安全风险，多国政府已宣布或考虑对其进行合理监管。

1.美国：审慎监管为主

美国对人工智能监管主要以审慎监管为主，监管力度相对宽松，重在激励与促进人工智能创新与发展。2020 年 1 月，美国联邦政府发布《人工智能应用监管指南》，提出针对人工智能技术和相关产业要以促进人工智能技术创新，减少技术应用障碍为宗旨而采取监管与非监管措施的指导建议。美国对人工智能监管呈现以下特点：一是对 ChatGPT 类生成式人工智能采取审慎监管，强调监管应有助于人工智能的创新与发展。为维

①房慧颖：《生成型人工智能的刑事风险与防治策略——以 ChatGPT 为例》，《南昌大学学报（人文社会科学版）》2023 年第 4 期。

②刘佳明：《人工智能在行政治理领域应用的挑战及对策》，《领导科学》2023 年第 5 期。

③张乐，童星：《人工智能的发展动力与分险生成：一个整合性逻辑框架》，《江西财经大学学报》2021 年第 5 期。

④张欣：《生成式人工智能的数据风险与治理路径》，《法律科学（西北政法大学学报）》2023 年第 5 期。

⑤郑世林，姚守宇，王春峰：《ChatGPT 新一代人工智能技术发展的经济和社会影响》，《产业经济评论》2023 年第 3 期。

⑥谢思，和军：《数字经济监管现状及变革研究》，《中国特色社会主义研究》2022 年第 3 期。

⑦Zhuo T Y, Huang Y J, Chen C Y, et al.“Exploring AI Ethics of ChatGPT: A Diagnostic Analysis”, arXiv preprint arXiv:2301.12867, 2023.

⑧陈兵：《促进生成式人工智能规范发展的法治考量及实践架构——兼评〈生成式人工智能服务管理暂行办法〉相关条款》，《中国应用法学》2023 年第 4 期。

护与保持美国在全球人工智能创新领域的领先地位,强调人工智能领域监管要以评估潜在监管措施对人工智能和创新发展的影响为决策依据,力争减少、移除人工智能技术发展和应用面临的不必要障碍,避免过度监管①。二是强调科学性与灵活性的软法规制。美国政府认为对人工智能监管与非监管举措要利用科学技术信息与相关流程,基于风险评估管理与成本效益分析,采用弹性的法律框架与灵活的监管构架,以适应人工智能程序更新。2023年美国发布的第一版《人工智能风险管理框架》属自愿适用的指导性文件,软法规制特征显著。为鼓励人工智能的创新与发展,美国政府更倾向于放松监管,提出对特定人工智能领域可以采取非监管措施,相关机构应考虑法律规定的豁免情形,允许为特定人工智能技术应用程序提供安全港的试点计划②。三是多方对话探寻治理框架。政府相关部门与业内人士对ChatGPT类人工智能的监管问题提出诸多建议。有"ChatGPT之父"之誉的山姆·阿尔特曼提出成立政府监管机构负责人工智能授权、引入许可证制度、建立安全标准、成立专家组对模型独立审计等监管建议。美国国家电信和信息管理局面向公众征集意见,以保证人工智能系统合法有效、合乎道德与安全问题。美国多家顶级人工智能公司相关负责人表示将为用户提供识别生成式人工智能的方法,并确保发布前对其安全性进行测试,同时在水印识别、安全性评估、加大网络安全投资、社会风险防控、前沿模型使用、第三方监管等诸多方面提出完善治理措施③。

2.欧盟:监管与竞争并举

欧盟人工智能监管表现为监管与竞争并举,率先搭建人工智能的立法与监管体系,监管相对更为严厉。为迎接人工智能的机遇与挑战,欧盟《人工智能白皮书》指出要大幅提高人工智能研究和创新领域投资,目标是未来10年内每年投入200亿欧元应用于人工智能开发。欧盟在人工智能治理方面的主要举措与特征表现:一是重视人工智能法律与治理体系的构建。欧盟委员会于2019年发布《人工智能伦理准则》,提出人工智能可信赖的七项标准;2020年发布人工智能数字战略系列文件:《通往卓越与信任的欧洲路径》《塑造欧洲的数字未来》和《欧洲数据战略》④;2023年欧盟委员会正式通过《人工智能法案(草案)》。此外,欧盟不仅出台《数字权利和原则宣言》,同时还颁布系列法案,如《数字服务法案》《数字市场法案》《数据治理法案》《数据法案》和《网络弹性法案》,并建立了欧盟数字身份框架计划、欧洲算法透明中心、GDPR认证体系以及数据跨境法律及执法判例研究等项目。二是创建多元化的监管模式与机制。欧盟人工智能监管与科技伦理监管体系的搭建具有显著的多元化特点。在监管主体方面,形成了立法、执行与行业协会不同层面的监管主体,其中欧盟委员会、欧洲议会、欧洲理事会共同参与人工智能科技伦理的立法与政策制定,同时设置了欧盟与成员国不同层面的执行监管机构,发挥了行业组织协会意见支持与促进性作用;在监管布局方面,欧盟强调以问题为出发点,突出人工智能重点领域重点问题的政策布局,特别是如何打造负责任与可信赖的人工智能的科技伦理监管等系列问题;在监管机制方面,欧盟成员国具有双重性监管特征,注重监管政策的综合性与政策细分,鼓励多元社会主体的共同参与⑤。三是人工智能立法与监管实践走在国际前列。2023年6月,欧洲议会全体会议通过《人工智能法案》授权草案,成为世界第一份关于规范人工智能的统一立法,展现出欧盟为人工智能设定全球标准的意图。该法案的突出特点是将人工智能风险划分为不可接受风险、高风险、有限风险、极小风险或无风险四个等级。其中,"不可接受风险"将严格禁止人工智能技术应用;"高风险"要求必须在严格监管条件下使用,程序研发者和使用用户要遵守数据管理、透明度、保存记录等相关规定,确保系统稳定、准确和安全

①史凤林,张志远:《论人工智能的公法规制:美欧模式与中国路径》,《理论月刊》2023年第8期。

②《从美国和欧盟的最新政策看人工智能的发展和监管》,全球技术地图,https://baijiahao.baidu.com/s?id=1711695197458097769&wfr=spider&for=pc。

③《AI"狂飙"引担忧 监管讨论在升温》,人民网,http://finance.people.com.cn/n1/2023/0704/c1004-40027303.html。

④《从美国和欧盟的最新政策看人工智能的发展和监管》,全球技术地图,https://baijiahao.baidu.com/s?id=1711695197458097769&wfr=spider&for=pc。

⑤肖红军,张丽丽,杨镇:《欧盟数字科技伦理监管:进展及启示》,《改革》2023年第7期。

（见表 1）。

表 1　欧盟人工智能风险等级分类与监管措施

风险等级分类	欧盟的监管措施
不可接受风险	禁止（被认为对人类安全与权力存在明显威胁的人工智能系统。如鼓励危险行为、社会评分等）
高风险	允许（须符合法律规定的合规内容并在投入使用前进行第三方评估后投入市场）
有限风险	允许（多数合规且存在风险的情形。对特定类别的人工智能系统设定透明度义务。）
极小风险或无风险	允许（鼓励制定促进人工智能系统自愿应用要求的行为准则）

注：根据欧盟《人工智能法案》内容整理。

3.日本：优先技术发展的软法监管范式

日本在人工智能监管特别是伦理监管方面已有了较为完整的理论与制度基础，其对人工智能技术发展与监管的关系问题形成了“既对立又统一”的基本立场。日本人工智能伦理监管以优先发展技术为前提，旨在为技术发展铺平道路，力争消除伦理问题产生的系列社会阻碍。相较于欧美国家，日本的人工智能伦理监管形成了以优先技术发展为主导的非约束性软法监管范式①。具体来看，日本对人工智能伦理监管举措有以下特点：一是建立多层次的制度架构。日本的人工智能伦理监管形成了不同层级的制度架构，包括原则层、规则层、监督层以及执法层。各层内容设计主要遵循政府伦理监管目标，实现目标路径的通用规则与特殊规则，引导企业自主规范的监督层，以及伦理监管的配套手段即对违规企业追责的执法层。目前，日本采取以软法监管为主，在执法层还比较薄弱，一定程度上影响了监管效果。二是搭建多元主体参与的监管运行体制。日本人工智能伦理监管主体包括政府、行业、企业以及社会等多元主体共同参与监管过程。政府以引导为主，通过民主讨论的方式凝聚共识，推动政策实施；行业协会与相关经济团体协同参与监管；企业主要提供技术信息，为促进人工智能技术发展提供具体要求，发挥自主规范与自我监督的积极作用；社会团体作为第四方力量，从不同视角与不同领域提出伦理问题与监管需求等。三是重视激励机制。日本在人工智能伦理监管的实践中，注重通过奖惩机制，特别是通过激励机制进行伦理监管。日本以表彰在业内领先的优秀企业以及将企业的反馈信息渗透于规则制定等方式，激发与增强参与企业自身的履职意识与责任感。与此同时，日本建立了认证制度。2022 年 4 月 1 日，开始实施“人工智能云服务的安全与可信赖信息公开认证制度”来审查企业，通过企业颁发证书与徽章，没有通过认证的企业一定程度上形成“惩劣”的效果。四是“硬法”监管用于特定领域。日本人工智能监管虽以软法为主，但在特定领域是存在硬法监管的。日本增加了在道路交通、医疗等领域的人工智能监管的法律补充与行政规定。2020 年 4 月日本修订《道路运输车辆法》和《道路交通法》，对无人驾驶等自动驾驶系统安全标准进行补充，明确使用主体责任与义务，对特定自动驾驶计划实施许可证制度。在医疗领域，2018 年补充《医师法》条款中指出使用人工智能技术治疗时，医生要对最终判断承担全部责任。此外，日本仍考虑将政府信息系统与行政服务等根据风险程度不同进行伦理监管②。

（二）国外主要经济体生成式人工智能监管经验启示

1.国外 ChatGPT 类生成式人工智能监管的基本经验

世界主要经济体在人工智能监管方面既有共性又有差异。共性特征主要体现在：一是国际社会与各国政府重视程度高。国际社会对 ChatGPT 类人工智能技术发展与治理问题普遍高度关注③。联合国呼吁采用人工智能“全球监管标准”，提出支持建立人工智能监管机构，计划于 2023 年底前启动高级别人工智能咨询机构工作，

①市川類.“AI 原則の体系化と今後のガバナンスの方向~デジタル・AIにおけるイノベーションと社会制度の共進化”，Institute of Innovation Research，Hitotsubashi University，2020，p.10.

②刘湘丽，肖红军：《软法范式的人工智能伦理监管：日本制度探析》，《现代日本经济》2023 年第 4 期。

③毛子骏，朱钰谦：《人工智能的国外社会科学研究热点综述》，《电子科技大学学报（社科版）》2023 年第 2 期。

定期审查人工智能监管进展。各国监管机构对人工智能监管也提出计划。英国国家相关监管机构已开始起草人工智能监管法规,并与多方协商以提高对人工智能技术的理解。G7领导人日本峰会着重探讨了生成式人工智能监管问题,计划起草"负责任的AI"标准并于2023年底出台对生成式人工智能的监管措施。澳大利亚是最早提出人工智能监管的国家之一,政府宣布成立"负责人的人工智能网络",同时监管机构拟对《隐私法》进行修改,进而完善人工智能监管的法律内容①。二是科技向善成为国际社会人工智能监管的新愿景②。各国对人工智能科技伦理监管十分关切,科技向善成为国际社会人工智能可持续发展与监管的关键特征。

各国在ChatGPT类人工智能监管上也具有明显差异。第一,监管原则不同。中国人工智能监管坚持包容审慎的基本原则,鼓励人工智能产业的健康持续发展;美国更倾向于放松规制,日本则以软法规制为主,而欧盟更侧重于强监管政策。第二,监管路径不同。中国人工智能监管的主要路径是通过制定并完善法律政策、规范标准来搭建监管机制,进而建立人工智能监管体系。相较中国,美国更倾向采取市场化的监管路径,日本则通过多元制度机制实施软法监管,而欧盟主要通过立法授权并鼓励企业自律合规。第三,监管侧重点不同。中国对人工智能监管更侧重于风险防控,关注生产内容审核、技术创新、服务规范和伦理与社会影响等方面。而美欧更侧重数据隐私、算法透明度等技术性防护。其中美国对风险结果的监管较为审慎,欧盟则倾向更为严厉的过程性规制③,而日本侧重在特殊领域进行硬性监管。第四,国际协调差异大。人工智能本身具有强跨境性,但目前各国在监管协调方面差异显著。中、美、欧都致力于打造国际监管标准,但因出发点与路径不同,实质进展有限。这也增加了跨境人工智能企业的监管复杂性,不利于技术跨境流通④。

2.国外ChatGPT类生成式人工智能监管的启示

通过比对国外人工智能监管的进展情况,可对我国人工智能监管事业的向前迈进有所启发。一是强化人工智能产业的战略性定位。人工智能产业的发展在推进中国式现代化进程中具有重要作用,要把人工智能产业作为重要的战略性新兴产业加以发展,增强其竞争力。二是结合"人工智能+"模式的市场化,提高监管的预见性和有效性。人工智能产业的发展潜力巨大,而政府监管的预见性和有效性无疑将对其可持续发展产生重大影响。三是加大政府支持。生成式人工智能技术具有使用成本高的特点,所以规范企业的行为离不开政府的鼓励与支持。四是加强国际协作,增强中国在全球人工智能产业的竞争力与话语权。一方面要加强国际间对话与协作,汲取技术与实践经验,取长补短,打造具有开放性和包容性的国际沟通环境;另一方面,贡献人工智能监管的中国方案,增强人工智能监管的国际话语权。

三　推进ChatGPT类生成式人工智能监管的中国方案

对ChatGPT类生成式人工智能的监管,要探求一个发展与安全并举、技术创新与社会进步双赢的中国治理方案,需要在合理借鉴国际人工智能监管实践经验的基础上,进一步紧密结合我国实际情况,从法律与政策、技术与安全、社会与伦理等层面提出监管对策。

(一)中国ChatGPT类生成式人工智能监管现状

中国对人工智能坚持以促进产业发展为基本导向的监管理念,并制定了一系列人工智能发展政策(见表2)。一是发展规划性政策。2017年,国务院发布了关于人工智能发展的顶层设计文件《新一代人工智能发展规划》,强调人工智能应遵循"以人为本、安全可控、公平公正、尊重隐私"的基本原则,提出建立健全人工智能治理体系。后又相继出台了《促进新一代人工智能产业发展三年行动计划(2018—2020)》和《国家新一代人工智能开放创新平台建设工作指引》等文件,以促进产业发展为导向,鼓励人工智能技术的研发和

①《AI"狂飙"引担忧 监管讨论在升温》,人民网,http://finance.people.com.cn/n1/2023/0704/c1004-40027303.html。

②《用"科技向善"理念引领人工智能发展》,人民日报海外版,http://www.cac.gov.cn/2019-05/06/c_1124454989.htm。

③史凤林,张志远:《论人工智能的公法规制:美欧模式与中国路径》,《理论月刊》2023年第8期。

④《从美国和欧盟的最新政策看人工智能的发展和监管》,全球技术地图,https://baijiahao.baidu.com/s?id=1711695197458097769&wfr=spider&for=pc。

使用。二是伦理规范性政策。2019 年，国家人工智能治理专业委员发布了《新一代人工智能治理原则——发展负责任的人工智能》，明确提出要遵循“和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作”的治理原则；2021 年发布了《新一代人工智能伦理规范》，提出增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养六项基本伦理要求，并对管理、研发、供应、使用等特定活动提出具体的伦理规范。2022 年中共中央办公厅、国务院办公厅印发《关于加强科技伦理治理的意见》的通知，强调科技伦理治理的总体要求，并提出了加快健全科技伦理治理体制、加强科技伦理治理制度保障、强化科技理论审查和监管以及深入开展科技伦理教育和宣传等意见。三是标准规范性政策。中国积极探索适应人工智能发展的法律框架，制定了专门针对人工智能的标准规范与法律法规，如 2020 年五部门印发《国家新一代人工智能标准体系建设指南》的通知；2023 年 6 月国家网信办发布《深度合成服务算法备案清单》；2023 年 7 月，国家互联网信息办联合国家发改委等 7 部门发布《生成式人工智能服务管理暂行办法》（以下简称《办法》），这是首个专门针对生成式人工智能监管的政策规范性文件，对中国人工智能监管具有里程碑意义。《办法》明确提出中国坚持发展与安全并重、促进创新与依法治理相结合的原则，内容涉及技术发展与治理、服务规范、监督检查与法律责任等，鼓励生成式人工智能创新发展①。

表 2　我国人工智能相关政策信息一览表

序号	政策文件名称	发文时间	发文机关	发文字号
1	《新一代人工智能发展规划》的通知	2017.7.20	国务院	国发〔2017〕35 号
2	《促进新一代人工智能产业发展三年行动计划（2018-2020 年）》的通知	2017.12.13	工信部	工信部科〔2017〕315 号
3	《国家新一代人工智能开放创新平台建设工作指引》的通知	2019.8.1	科技部	国科发高〔2019〕265 号
4	《新一代人工智能治理原则——发展负责任的人工智能》	2019.6.17	国家新一代人工智能治理专业委员会	——
5	《国家新一代人工智能标准体系建设指南》的通知	2020.7.27	标准委、网信办、改革委、科技部、工信部	国标委联〔2020〕35 号
6	《新一代人工智能伦理规范》	2021.9.25	国家新一代人工智能治理专业委员会	——
7	《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》的通知	2022.7.29	科技部、教育部、工信部 、交运部、农业农村部、卫健委	国科发规〔2022〕199 号
8	《关于支持建设新一代人工智能示范应用场景》的通知	2022.8.12	科技部	国科发规〔2022〕228 号
9	《关于加强科技伦理治理的意见》的通知	2022.3.20	中办、国办	中办发〔2022〕19 号
10	《生成式人工智能服务管理暂行办法》	2023.7.13	国家网信办、国家发展改革委、教育部、科技部、工业信息化部、公安部、广电总局	——

资料来源：中国政府网、中华人民共和国科学技术部等官方网站，整理编制。

中国人工智能监管已取得一定进展，初步建立了人工智能监管的基本架构且正在有序推进。在战略上，中国政府已将人工智能列入国家战略性新兴产业，明确要加快人工智能产业的发展；在资金投入上，政府出台了系列支持政策，如设立专项资金、引导社会资本投入等；在人才建设上，通过人才引进政策吸引国内外优秀人才进入人工智能领域进行研究与应用实践；在产业发展规划上，

①《生成式人工智能服务管理暂行办法》，中华人民共和国中央人民政府网，https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm。

发布多项产业发展规划性政策,明确了人工智能产业的发展目标;在扶持政策上,在税收优惠、知识产权保护和科技成果转化等方面给予诸多扶持,为人工智能发展提供环境与政策支持。此外,政府相关部门积极有效推进政策的制定与落地助力人工智能产业快速发展。自2017年新一代人工智能发展以来,政府有关部门相继发布多项人工智能发展规划与专项政策。特别是2022年ChatGPT出现后,相关部门发布了《生成式人工智能服务管理暂行办法》,充分体现了有关部门积极推动生成式人工智能政策的制定与完善。

(二)中国ChatGPT类生成式人工智能的监管对策

1.基于法律与政策层面的监管对策

(1)强化法律衔接,推进综合性立法

现行法律规范与政策滞后,存在对人工智能复杂场景难以监管的问题。对此,提出以下监管对策:第一,针对生成式人工智能领域进行顶层综合性立法,明确主体责任,建立问责机制。目前,我国尚未出台专门规范人工智能的顶层综合性法律,对人工智能产业发展中涉及的基本风险与各方主体权责划分问题缺乏监管,阻碍了人工智能技术的拓展应用。因此,要加快完善人工智能领域的专项法律法规,对人工智能的潜在风险建立问责机制,规范主体责任,明确研发者、运营者、使用者等多方权责并落实问责程序。第二,修订完善相关法律法规细则。我国虽已出台诸多人工智能监管政策,但监管政策仍落后于技术发展。要加快制定生成式人工智能的责任细则,并通过补充调整相关条款内容以适应人工智能技术的发展,推进生成式人工智能的法治建设和治理体系的持续完善。

(2)重视法治数字化建设,强化监管政策落地

生成式人工智能技术革新与迭代速度快致使其监管不同于传统产业的监管特征,这对监管技术的数字化建设提出了新的要求。法律是生成式人工智能监管的制度性保障,法治监管的数字化建设与技术革新发展则是人工智能监管政策落地的重要支撑。因此,重视法治监管与数字技术的融合发展是不断推进生成式人工智能监管政策实施与落地的关键环节。而如何将人工智能监管法律法规与规制细则进行代码转化,完善算法程序的法治技术转化①,是目前生成式人工智能监管数字化亟需解决的重要内容。通过法治数字化的建设与发展,一定程度上消除了法律规则的人为解释与操作的不确定性,提升了生成式人工智能监管的客观性与可信度,进而提升政府监管的数字化能力。

(3)建立行业标准规范,明确应用场景范围

ChatGPT类生成式人工智能具有广阔应用场景,但在不同行业或场景应用风险也存在较大差异。因此人工智能监管需要行业细分,对不同应用场景进行风险评估。例如,建立分类别生成式人工智能应用风险评估标准与评估机制,制定不同场景和产业审核标准。对此,需要政府及相关部门发挥引导作用,加快人工智能监管实践的步伐。同时要以需求为导向,依据人工智能不同发展阶段及其技术特征,找出相应风险挑战与问题并制定有效政策。建立政策评估与监督机制,补充法律漏洞,灵活应对人工智能技术革新变化与挑战。

2.基于技术和安全层面的监管与对策

(1)建立进入性监管,做好事前合规评估

人工智能在应用之前,应从源头上开展进入性监管,做好事前监管。具体上看,一是对研发主体资质与产品进行合规评估,通过准入审核与弹性审批相结合的方式进行监管。在人工智能研发初期,需结合人工智能企业开发的产品与服务特点进行风险评估,此阶段应重点关注研发主体资质及其行为审核,通过准入制度与弹性审批相结合的方式以避免过度监管。目前,深圳与上海对高风险人工智能的应用设置了事前合规评估与审查,而对中低风险人工智能应用则采用事先披露与事后控制相结合的监管模式。二是算法风险的事前评估与审查。对生成式人工智能技术的核心算法需要求研发主体接受算法审查,可采取审查公开源代码与系统程序、监测研发主体操作行为等静态与动态相结合的监管方式,以便及时纠偏

①陈兵:《促进生成式人工智能规范发展的法治考量及实践架构——兼评〈生成式人工智能服务管理暂行办法〉相关条款》,《中国应用法学》2023年第4期。

以确保程序运行的正向稳定①。

（2）强化技术性监管，落实全程动态监管

第一，对数据安全风险的合规处置。对国家数据安全建立分级审查监管机制，保护数据主权；对政务数据，采取先报备后公开并设立有限加工利用的先决条件；对个人数据，设置监督机构审查数据搜集深度以防范虚假回复。在技术安全防控方面，需加快技术防护建设，组织安全技术专家、安全技术企业尽快提出相应的技术方案。

第二，对算法模型的纠偏与监管。为避免生成式人工智能在算法设计、数据训练、模型生成等过程出现歧视性问题，应加强技术性监管，特别是算法共谋监管，做好全流程监管。一方面，针对算法模型偏见，可通过调整算法学习路径，对固有算法程序进行预防、调整与校对，规避偏向性误导；同时研发者应对生成式人工智能核心算法的透明度和解释性方面提供技术支持，在保护用户数据隐私、消除模型偏见的同时，应确保问责机制的设立与运行，评估模型潜在的负向影响。另一方面，要加强算法共谋监管。数字经济市场中，人工智能算法的应用使得市场共谋表现得更为隐蔽和便捷。在不同的算法作用下，可达成不同形式的算法共谋，从而影响市场公平与竞争。由此需建立动态监管机制，实时监管网络生态，加快技术性监管突破，落实事前、事中、事后全流程动态监管，进而遏制算法共谋②。

（3）加快行业风险评估，建立分类分级监管

为避免过度监管，基于风险分析的人工智能系统分类分级监管具有良好的实践前景。人工智能作为未来世界科技竞争的关键，在保持国际竞争优势前提下，各行业对人工智能风险评估具有重要的意义。如何评估并将某类人工智能系统归类为不可接受风险、高风险、中等风险、低风险或无风险等不同风险等级，成为监管实践的重点与难点。对此，需要加快搭建政府、行业、企业等多方参与的沟通平台，加快建立人工智能领域不同应用场景下的风险评估标准与评估机制，推动行业与企业自我监管。

3.基于社会和伦理层面的监管与对策

（1）加快科技伦理建设，打造负责任 AI

人工智能系统研发需强调社会伦理价值，研发者要充分考虑人工智能产品是否符合人类社会伦理规范和道德准则。对此，ChatGPT 类人工智能产品在投入市场前需完成相关法律法规的预训练学习，细化完善科技伦理监管框架和制度规范，通过科技伦理标准的设定、执行与检测检验等事前监管方式使智能机器的自主决策行为遵守人类社会法律规范并符合伦理价值观，生成符合社会伦理价值的“负责任”人工智能系统。具体上，人工智能在进行行为决策过程中要避免歧视偏见以及侵犯用户隐私等不当行为，对可能引发的社会伦理问题进行有效预防与纠正。由于人工智能系统伦理考量是一个持续过程，因此研发者应定期对人工智能系统决策行为进行持续伦理审查与改进，对潜在伦理风险进行及时评估与防范，确保生成式人工智能系统与社会范畴伦理价值观的一致性。

（2）加强协同监管，促进行业自律

第一，加强政府部门间协同监管。政府在生成式人工智能产业的发展过程中发挥着主导作用，要积极搭建政府与企业、社会相关组织、公众多方参与的协同监管平台。由于 ChatGPT 类生成式人工智能涉及到广泛的应用场景，因此在政府监管实践中促进部门间的沟通与协调尤为重要，有利于部门间的信息共享与高效协作，实现综合治理。第二，引导并促进行业自律。行业自律是人工智能监管不断向前发展的重要基础，而法律法规与标准规范的建立与完善客观上成为有利于促进行业自律的外部条件。第三，鼓励政府、企业、学界与社会组织的多方参与对话。生成式人工智能监管需要多方合作促进产业创新与技术发展，通过多方参与合作与对话促进政策制定与完善，确保政策内容能考虑不同利益主体需求，既有利于政策实施效果，同时有利于促进企业的社会责任行为与产业可持续发展。

（3）协调创新与监管，兼顾发展与安全

ChatGPT 类生成式人工智能的中国方案需协调创新发展与监管平衡，兼顾发展与安全。第一，把握好监管强度。过度监管，特别是在内容审核、算法备案等方面的过度要求，势必增加企业负担

①刘佳明：《人工智能在行政治理领域应用的挑战及对策》，《领导科学》2023 年第 5 期。

②和军，谢思：《政府监管研究进展与热点前沿（2021—2022）》，经济科学出版社 2023 年版，第 150—151 页。

并制约技术创新,不利于产业发展。因此,政府监管要在过度监管与放松监管间找到平衡。第二,提高监管规则透明度。监管规则的不透明与频繁变化会增加产业不确定性,进而影响企业中长期规划并制约产业发展。因此,监管部门应提高决策透明度,在制定监管政策前充分论证、征求意见。第三,增强监管内容的精准性。人工智能监管内容与要求的不明确容易导致监管效果不显著甚至失灵,增添企业成本,影响企业创新发展积极性。因此,需根据不同应用场景或行业领域进行精准施策。第四,提高监管政策的适应性。ChatGPT 类人工智能技术的迭代速度要求监管政策具有良好适应性,能灵活调整适应技术变化需求。若监管机制无法与时俱进,必然无法达到预期监管效果。第五,增强国际间治理的协同性。ChatGPT 类人工智能产业的国际化特征决定了不同国家的监管政策会相互影响。加强国际协作,推动国际人工智能标准设定,助力我国人工智能产业的蓬勃成长,提高在国际社会上人工智能治理中国方案的竞争力与话语权,进一步完善人工智能治理体系。

综上所述,ChatGPT 类生成式人工智能监管具有交叉性特点,不同场景具有多种风险挑战。例如,数据合规问题,不仅涉及法律层面的监管不完备及技术层面的事前监管与进入性监管问题,也涉及数据隐私方面的社会伦理问题。此外,ChatGPT 类生成式人工智能的监管需关注全产业链监管方案,完善对研发初期、进入期、运营期以及内容生成后的全链条监管机制,做好事前、事中、事后的全过程监管。最后,人工智能监管以服务与促进人工智能产业发展为宗旨,坚持发展与安全并重、促进创新与依法治理相结合的监管原则,为打造“可信、可控、向善”的生成式人工智能提供制度性保障,实现技术发展与社会进步双赢,助力推进人工智能治理体系的持续完善。

International Comparison and Reference on ChatGPT-like Generative Artificial Intelligence Regulation

HE Jun & YANG Hui

(School of Economics, Liaoning University, Shenyang 110036, China)

Abstract: The rapid development of ChatGPT-like generative artificial intelligence has brought positive impacts to different industries and fields. At the same time, it has led to a series of risk challenges and regulatory issues at the legal and policy, safety and technology, social and ethical aspects. This paper provides inspiration and necessary reference for the further improvement of China's artificial intelligence governance system by refining and comparing the artificial intelligence supervision and governance practices of major international economies, such as the United States, the European Union, and Japan. This paper proposes to improve laws and regulations, industry standards and related regulatory policies, establish entry regulations, technical full process regulations, hierarchical and classified regulations, strengthen technology ethics, multi-party collaborative regulations, coordinate innovation and supervision, and other regulatory measures to further promote the healthy and sustainable development of artificial intelligence and improve the artificial intelligence governance system.

Key words: ChatGPT; generative artificial intelligence; risk challenges; international practice; China's path

(责任校对 朱春花)