

生成式人工智能的智慧监管： 理论、模式与选择*

张树华 张夏添

【摘 要】生成式人工智能作为极具影响力与颠覆性的新兴技术和新质生产力的核心要件，正以新理念、新业态、新模式深刻影响与改变人类经济、政治、文化、社会的诸多领域。这一技术既是新一轮科技革命与产业变革的重要驱动力量，又在快速迭代与突破中为监管体系带来新挑战与新要求。针对生成式人工智能的要素层、技术层、内容层、产业层、发展层的五维风险，文章旨在智慧监管的理论脉络与参考依据下，对比分析国际现有监管方案，尝试提出建构统筹发展和安全、效率和公平、活力和秩序的智慧监管模式，从而保障与促进生成式人工智能的安全、有序、高质量发展。

【关键词】生成式人工智能 智慧监管 监管理论 人工智能治理

【作者简介】张树华，中国社会科学院政治学研究所所长，中国社会科学院大学政府管理学院院长、教授、博士生导师；张夏添，中国社会科学院大学政府管理学院博士研究生。

【中图分类号】D922.17 **【文献标识码】**A

【文章编号】2097-1125（2025）05-0026-25

* 本文系中国社会科学院学科建设“登峰战略”资助计划“中外政治与国家治理”（DF2023YS7）、国家社会科学基金重点项目“健全全过程人民民主制度体系的理论深化与实践创新研究”（24AZD070）的阶段性成果。

一、引言

当前，世界百年未有之大变局全方位、深层次加速演进，以人工智能为代表的新兴技术正与经济社会发展和人类生活重大需求深度融合，并进一步影响和形塑全球发展格局。习近平总书记深刻指出：“人工智能是引领这一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的‘头雁’效应……正在对经济发展、社会进步、国际政治经济格局等方面产生重大而深远的影响。加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手，是推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源。”^① 面对新一轮科技革命和产业变革，2024年7月18日，党的第二十届中央委员会第三次全体会议通过《中共中央关于进一步全面深化改革 推进中国式现代化的决定》，进一步指出要“健全因地制宜发展新质生产力体制机制。推动技术革命性突破、生产要素创新性配置、产业深度转型升级，推动劳动者、劳动资料、劳动对象优化组合和更新跃升，催生新产业、新模式、新动能，发展以高技术、高效能、高质量为特征的生产力……完善推动新一代信息技术、人工智能、航空航天、新能源、新材料、高端装备、生物医药、量子科技等战略性新兴产业发展和治理体系，引导新兴产业健康有序发展”。^②

生成式人工智能是人工智能技术的最新突破，是发展新质生产力并推动高质量发展的核心所在。同时，生成式人工智能作为人类社会遇到的第一种将多类技术特性相融合的复合型技术，它的快速迭代突破与融合交互使得相应的系统性风险不容小觑。^③ 当前，全球各国及国际组织已就生成式人工智能的安全与发展提出各具特色的监管方案，较有代表性的包括我国出台的《生成式人工智能服务管理暂行办法》（以下简称《办法》）、欧盟颁布的《人

① 《习近平在中共中央政治局第九次集体学习时强调 加强领导做好规划明确任务夯实基础 推动我国新一代人工智能健康发展》，《人民日报》2018年11月1日。

② 《中共中央关于进一步全面深化改革 推进中国式现代化的决定》，人民出版社2024年版，第10~11页。

③ 参见Allison Stanger, Jakub Kraus and Woojin Lim, et al., Terra Incognita: The Governance of Artificial Intelligence in Global Perspective, *Annual Review of Political Science*, Vol. 27, 2024, pp. 445–465。

工智能法案》(以下简称《法案》)、英国推行的《支持创新的人工智能监管方式白皮书》(以下简称《白皮书》)及美国的系列监管政策。本文旨在基于生成式人工智能的多重风险,在智慧监管理论的背景下提出未来对生成式人工智能进行智慧监管的路径与建议。

二、生成式人工智能的特点与风险

人工智能生成内容(AI-Generated Content, AIGC)指通过人工智能算法对数据进行收集、筛选、标注、分析、训练、加工、再创造而智能生成的内容,包括音频、文本、图像、视频及多种混合模态。生成式人工智能是人工智能技术、内容、应用、行业等多个维度的最新迭代,既是科技层面的重大创新和变革,也以高度灵敏、深度交互、快速迭代、广泛场景等强大的技术扩展能力与智能跃迁能力而深度渗透和形塑诸多行业,成为全球科技、经济、文化、政治等领域的新型基础设施。

生成式人工智能既已取得迅猛发展,又具备未来创新突破的巨大空间。^①自2022年11月30日OpenAI公司推出生成式人工智能的标志性产品ChatGPT以来,这一跨时代革命性的新技术以其令人惊艳的智能水平与颠覆性的变革能量掀起全球生成式人工智能创新热潮,已成为全球经济繁荣的主要驱动力,并将进一步影响甚至重构国际科技战略布局与发展格局。当前,全球各国及国际组织已快速入局和全面布局以生成式人工智能为代表的新兴技术领域,并将持续加大对其资金及政策支持。美国、英国、日本、北约、欧盟等国家和国际组织已将生成式人工智能定义为关乎国家发展命运与全球权力博弈的核心技术,指出生成式人工智能将是大国竞争最为重要的领域。在高速增长和快速发展之下,新兴技术的新颖性、不确定性、模糊性会对经济社会产生显著影响,并在要素层、技术层、内容层、产业层及全球网络安全领域带来前所未有的冲击和挑战。

一是要素层风险。生成式人工智能技术的核心要素包括数据、算法和算力。在数据要素上,生成式人工智能技术的研发过程包括但不限于原始

^① 参见 Manyika, James and Michael Spence, *The Coming AI Economic Revolution: Can Artificial Intelligence Reverse the Productivity Slowdown?*, *Foreign Affairs*, Vol. 102(6), 2023, pp. 70-86。

数据集中收集、人工或智能数据标注、数据筛选检测、数据清洗和再次处理、模型训练和迭代优化等诸多环节，数据已不只是生产资料，更可能在单个环节和多个步骤交叉互动中造成数据泄露等安全问题。^① 在算法要素上，算法偏见、算法歧视、算法黑箱导致的带有偏见乃至意识形态色彩的生成式人工智能将干扰、误导乃至再塑造国际交流互动，信息茧房将进一步加剧偏见和扩大隔阂，产生新型信息贫困和数据鸿沟。^② 在算力要素上，由芯片和电力驱动的算力消耗随着数据的爆发和训练规模的增大而呈现指数级增长，这不仅在能源领域掀起环保和资源危机，更由于芯片的稀缺性与重要性而使其愈发成为全球科技及贸易博弈的最新标的，在算力治理领域呈现新的全球竞争态势。^③

二是技术层风险。目前对生成式人工智能技术及其具备的智能跃迁等先进能力的解释与研究仍有不足，超大规模的训练模型内部运作机制及其未来智能水平仍充满不确定性，生成式人工智能技术将如何演进乃至如何进化尚是未知数。由于生成式人工智能可以更为迅速且智能地抓取、吸收、学习乃至再创造目前已知的几乎所有信息，并使用巨大的计算资源进行海量复制和迭代，故而在此前所未有的自我学习的智能背景下，技术上的不可知与不可靠或如“潘多拉魔盒”一般释放人类难以预测和把控的巨大风险。从技术系统看，生成式人工智能的超大训练规模并不能完全提升其鲁棒性，^④ 甚至会在某些方面增加出现程序崩溃、系统坍塌等严重后果的可能性。^⑤ 随着生成式人工智能技术逐渐深度嵌套进金融、医疗、军工等多个领域，技术的危害可能被连锁性放大。

三是内容层风险。生成式人工智能技术实现了音频、视频、文本等多模

-
- ① 参见张文魁：《数据治理的底层逻辑与基础构架》，《新视野》2023年第6期，第63~71页。
- ② 参见钟祥铭、方兴东：《智能鸿沟：数字鸿沟范式转变》，《现代传播（中国传媒大学学报）》2022年第4期，第133~142页。
- ③ 参见周辉、闫文光：《中国人工智能算力产业发展的规制困境及其解决路径》，《北京工业大学学报》（社会科学版）2025年第1期，第87~102页。
- ④ 鲁棒性（robustness）指程序在非正常情况下的系统生存能力，在计算机科学和软件工程领域通常指系统在恶意攻击、恶性干扰等异常情况下仍旧保持正常运行和准确输出的能力。
- ⑤ 参见陈炫婷、叶俊杰、祖璨等：《GPT系列大语言模型在自然语言处理任务中的鲁棒性》，《计算机研究与发展》2024年第5期，第1128~1142页。

态内容的自动生成。生成式人工智能技术赋予机器学习和吸收人类知识的能力，使其具备自我生产和智能再造碳基生物熟悉的信息结构的智慧水平，又反过来再次被多元主体使用与利用，完成人类社会内容生产的大爆发。^①一方面，基于这种低门槛、低成本、高质量、快反应的生成式内容已遍布网络，虚假音频、图像、视频将对大众传播造成前所未有的内容冲击。另一方面，基于生成式人工智能技术而生成的内容的知识产权等议题也对既往知识体系形成挑战甚至带来颠覆性影响，对相关问题的识别、界定、认证、处理尚处灰色地带。

四是产业层风险。生成式人工智能不仅是一项新兴人工智能技术，而且包含以此技术为发展基础与核心驱动的诸多产业。一方面，全球范围内主导生成式人工智能创新发展的科技公司股权结构复杂、发展渠道多样，部分国家的科技巨头已拥有超越、制衡甚至对抗国家主权的巨大能量。同时，生成式人工智能技术已被快速并广泛地应用于文化交流、国际传播、金融贸易、生物医药、自动驾驶甚至核能等领域，与生成式人工智能技术相关联产业的不可控性与不安全性被再度放大。^②在该发展特性下，开发者与监管者之间的技术壁垒塑造了产业发展的灰色地带，在生成式人工智能产业与其他领域复杂多维的交互中造成进一步的治理难题。另一方面，以生成式人工智能为代表的新兴产业具备“一管就死，一放就乱”的发展特点，^③在大国技术博弈背景下，引导科技向善和有效促进产业发展往往难以平衡和兼顾。

五是全球网络空间安全风险。本轮以生成式人工智能等新兴技术主导的权力争夺与再分配是全球秩序博弈的最新发展，技术已成为撬动全球权力分布的绝佳杠杆。生成式人工智能技术的快速发展打破了现有国际科技力量的微妙平衡，将多角度、全方位地改变人类社会运行的底层逻辑与发展脉络，在削弱当前全球战略稳定性的同时可能激发新的全球性危机。一方面，生成式人工智能时代网络犯罪、网络攻击等问题已产生质变，大国网络空间安

① 参见张夏添：《生成式人工智能技术与国际传播新格局》，《世界社会主义研究》2024年第9期，第112~124页。

② 参见 Daniele Rotolo, Diana Hicks and Ben R. Martin, What Is an Emerging Technology?, *Research Policy*, Vol. 44(10), 2015, pp. 1827-1843。

③ 参见薛澜、赵静：《走向敏捷治理：新兴产业发展与监管模式探究》，《中国行政管理》2019年第8期，第28~34页。

全具备辩证性、历史性与策略性。^①另一方面，从网络安全到网络安全化，安全一词已成为部分国家入侵他国数据要素、长臂管辖他国技术发展、攫取地缘政治利益、夺取全球霸权的战略性工具。技术民主联盟等围绕技术资源的斗争此起彼伏，部分人工智能创新大国正加速解构既往技术治理合作框架，着力打造排他性闭环产业链价值链，以身份政治和丛林法则等逻辑施行脱钩断链，筑起“小院高墙”，不断加剧全球技术权力之争。^②全球网络安全形势不仅日益复杂，而且在逐步走向对抗与分裂。在生成式人工智能技术的快速更迭下，全球性技术交流与合作、监管与治理的困难骤然加剧。

为更好地迎接挑战、应对风险、促进发展，对生成式人工智能展开有效监管已迫在眉睫，不少学者就何以监管生成式人工智能展开探讨。较多学者从法律层面提出生成式人工智能的法律规制手段，^③并设计风险化解方案，^④部分学者从治理角度出发探讨生成式人工智能的发展之道。新兴技术兼具收益和风险的二重性，相关监管治理目标不仅需要包括建构促进技术发展的规则，而且要基于技术特性实现风险预防和协调发展，找到更为科学有效的监管之策。

三、智慧监管理论的嬗变与演进

监管存在于个人、市场、政府、社会等多元主体的交流、互动、合作、博弈乃至对抗的复合过程中，监管理论的建构反映不同文化背景与惯例遵循下各国的发展现实和道路选择。监管理论随着实践推进而嬗变，监管实践在理论创新指导下走深走实，形成双向促进循环。

-
- ① 参见孔亦思、李抒音：《大国网络空间安全博弈：一种复杂系统分析视角》，《俄罗斯东欧中亚研究》2025年第1期，第41~60页。
- ② 参见李应博、林子秋：《技术权力与大国创新：西方国家认知逻辑、行为模式与外溢后果》，《政治学研究》2023年第5期，第126~142页。
- ③ 参见刘艳红：《生成式人工智能的三大安全风险及法律规制——以ChatGPT为例》，《东方法学》2023年第4期，第29~43页。
- ④ 参见毕文轩：《生成式人工智能的风险规制困境及其化解：以ChatGPT的规制为视角》，《比较法研究》2023年第3期，第155~172页。

（一）监管理论的创新与发展

监管亦或是管制、规制、监督管理等对“Regulation”元概念的阐释，在本质属性、因循惯例、功能作用等定义方式及时空选取的不同下存在一些差异。早在1987年，英国学者罗伯特·鲍德温（Robert Baldwin）等在梳理西方学界关于监管概念探讨的基础上，根据监管主体、监管范围、监管对象的具体差异，将监管模式分为三种显著不同的类型。^①2002年，英国学者茱莉娅·布莱克（Julia Black）对监管做出进一步阐释，指出监管理论应包含何为监管、监管主体、监管对象、监管方式等维度。^②2017年，英国学者克里斯特·库普（Christel Koop）等提出监管概念及理论应包含和回答的五个关键问题，即监管是否仅限有意识的行为、监管是否只能由政府实施、监管的干预手段和工具有哪些、监管范围是否局限于经济活动、监管过程是否需要区分主体和对象。^③

在早期的监管理论中，以英美为首的国家具备经济快速发展、市场风险迅速增长以及国家监管活动快速扩张与壮大的特点，基于此形成的以政府与市场的二元关系为出发点而建构的传统政府规制型的监管理论认为，政府可以作为监管的唯一主体并充分且集中地使用所有监管权，通过政府的强制力量来实施管制措施。^④20世纪末，市场经济的蓬勃壮大使原有的政府规制型监管高成本、低效率等问题凸显，监管主体多元化越来越被学界及业界接受。区别于新古典经济学派强调的完全市场、完全信息、完全独立和国家干预学派指出的导致市场失灵的国家干预因素，在自由主义、多元主义、社团主义、私益理论等主张的蓬勃发展下，探索平衡政府及市场关系的多元主体协同共管型监管理论开始兴盛。学者们逐渐认识到，监管制度不会一成不变，既要有意识地设计监管制度，又要以系统观念从更高维度关注经济、政治、法律系统的协作，考虑多元主体在复杂现实中的互动以及理论和实践的塑造与再

① 参见 Robert Baldwin and Christopher McCrudden, eds., *Regulation and Public Law*, London: Weidenfeld and Nicolson, 1987, pp. 45-60。

② 参见 Julia Black, *Critical Reflections on Regulation*, *Australian Journal of Legal Philosophy*, Vol. 27, 2002, pp. 1-35。

③ 参见 Christel Koop and Martin Lodge, *What Is Regulation? An Interdisciplinary Concept Analysis*, *Regulation and Governance*, Vol. 11(1), 2017, pp. 95-108。

④ 参见徐国冲、黄丽妹：《政府监管的演变历程与逻辑动因：从政府与市场关系变迁说起》，《东南学术》2025年第2期，第156-168页。

塑造。

理论界相继提出了去中心化监管（decentered regulation）、回应性监管（responsive regulation）、自愿性监管（voluntary regulation）、协同性监管（collaborative regulation）、自我监管（self regulation）、监管多元主义（regulatory pluralism）等理论和概念。其中，应用较为广泛且产生较大影响的莫过于回应性监管理论。1992年，伊恩·艾尔斯（Ian Ayres）等提出回应性监管理论，在纵向上建构按照层级划分的监管策略并配套差异化的监管处罚，在横向上以不对称监管、强化型自我监管、三方理论对监管权进行顶层设计。回应性监管分别对个体企业和行业设计金字塔式的执行策略，包含劝说、警告、民事处罚、刑事处罚、执照暂停及吊销执照等涉及从自我监管、强制性的自我监管到实行自由或非自由裁量惩罚的命令式监管的一系列执法执裁。^①随后，不同研究者将回应拓展至对监管体制环境、监管工具逻辑等变化的反应，并指出需要超越基于单一国家、单一主体、单一策略的传统回应性监管，对多样动机的多种行为者设计更为综合的回应性监管，构建刚柔并济的监管策略。^②同时，数智技术的演进驱动新型监管秩序的发展与建构。^③

监管主体多元化是监管实践和监管理论演进的主要方向，以人工智能技术为代表的新兴技术的应用与扩散极大增强了非政府主体的监管能力，而且监管体制日益呈现去中心化的特征。一方面，监管者与被监管者之间由于技术与专业知识隔阂导致的信息不对称使以政府为核心的监管主体难以掌握行业及市场的发展规律与特点。另一方面，政府、市场、公众等不同主体之间的权利壁垒在逐渐消解，监管日益成为三向甚至多向的互动，^④监管与风险的互动关系驱动监管理论的不断发展。^⑤基于国家中心主义前提而提出的回应性监管理论在制度设计、个体能动性、整体性干预等方面存在一定的

① 参见 Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate*, Oxford: Oxford University Press, 1992, pp. 35–52。

② 参见任梅、代栋栋：《生成式人工智能的回应性监管架构及其实现路径》，《中国行政管理》2025年第2期，第39~50页。

③ 参见史安斌、俞雅芸：《国家在场：迈向全球数智技术的新监管秩序——基于2024年互联网治理研究的主题聚类分析》，《编辑之友》2025年第2期，第45~55页。

④ 参见刘鹏、李杰钊：《新时代的中国信用监管：理论维度、模式分析和实践应用》，《新视野》2024年第5期，第89~98页。

⑤ 参见 Julia Black and Robert Baldwin, When Risk-based Regulation Aims Low: A Strategic Framework, *Regulation and Governance*, Vol. 6(2), 2012, pp. 131–148。

局限性。在超越“国家—企业”的二元主体框架和回应性监管理论建构的三方理论的监管模型基础上，智慧监管（smart regulation）逐渐被广泛应用和推广。

（二）智慧监管的理论特征与演进

1998年，尼尔·甘宁汉（Neil Gunningham）等提出了智慧监管理论。^①在过往理论的基础模型上，智慧监管理论丰富了基于不同主体的监管工具、协调了监管体制中的不同运作机制、增强了在不同情境下监管策略的适应性。

在监管理念上，智慧监管增强了监管网络和监管空间的系统性与协调性，主张以“大监管”理念实施包容审慎的监管策略。在监管主体上，智慧监管在监管多元主体的基础上重视政府与被监管方、第三方等不同主体的合作监管。在监管工具上，智慧监管为政府、作为自我监管的企业、以公共利益集团等为代表的第三方设计了混合型监管工具，三方可使用的监管工具分布依照回应性监管的金字塔型呈现由疏到密、由松到严的监管强度。其中，政府可以采取建议、教育、说服、警告、通知、刑事制裁、取消资格或执照等手段进行监管，作为自我监管的企业和第三方可以采用培训监督、激励、审查、晋升等多元工具。同时，信息化、数字化的监管工具被广泛吸收和用于智慧监管的混合工具策略中，通过激励性、限制性、惩罚性导向的不同手段增强监管工具的匹配性与恰当性，针对不同主体的自身特点与优势量身定做可供取用的监管手段。在监管效果上，智慧监管主张在复杂动态的监管环境下采取多种手段降低监管成本，提升监管效率，平衡好安全与发展的关系。

需要注意的是，智慧监管虽然在监管工具上仍基本遵循金字塔型的工具设计，但是相较回应性监管或基于风险的监管模式，^②智慧监管的核心特点在于政府的首要角色是创造必要的监管前提条件，使第二或第三方能够承担更多的监管责任，而不是直接大幅度介入整体性的监管方案中。一方面，政府作为引导者、协调者、掌舵者，应尽量减少对有限监管资源的消耗，并使社

① 参见 Neil Gunningham, Peter Grabosky and Darren Sinclair, *Smart Regulation: Designing Environmental Policy*, Oxford: Oxford University Press, 1998, pp.37-42。

② 参见刘鹏、张崙楠、王力：《基于风险的政府监管：理论发展与实践应用》，《中国行政管理》2024年第3期，第111~124页。

会层面直接关心监管问题的更多主体能够有效参与到监管进程中。依托智能算法等建构的监管体系也存在一定程度的决策风险。^①另一方面，政府可以在一个综合性的工具金字塔中促进监管方式的协调和升级，填补可能存在的监管空白，并增强其不同层次之间的联系。

智慧监管既是从监管理念、监管主体、监管工具、监管效果等方面对传统政府规制型监管及回应性监管等监管模式的超越与延伸，又是借助新技术新手段在生成式人工智能技术快速爆发和突破背景下智慧化的新型监管。加拿大政府指出，智慧监管对监管对象而言既是保护又是促进，它的核心在于尽可能提高监管的有效性，确保监管不会过于复杂或昂贵。同时，智慧监管能够自我更新，并跟上科学、技术和全球市场的发展，这意味着在保持高标准并采取适当问责措施的基础上，给予被监管者更多的灵活性。2024年5月，英国政府更新了新型智慧监管框架，指出智慧监管是只在必要时使用监管，并确保其设计和使用既适当又具有未来性。智慧监管具有三大特征：一是改革现有法规，以减轻监管负担并确保法规是当代的和具有前瞻性的；二是将监管作为最后的手段，而不是首选，这包括在有利时尽量利用替代的监管手段；三是确保良好运作的监管格局。

随着大数据、云计算、人工智能技术的兴起与运用，智慧监管已不局限于对原有理论的阐释与实践，智能社会的复杂性提出了新的监管难题。^②“智慧”一词在新的技术条件下也被赋予包含监管智能化、数字化、科技化的意指。有学者指出，在数字经济时代，智慧化、数据化、平台化、动态性是监管转型的重要特征，政府监管应基于整体治理视角建构包含内容、载体、机制、监督和评价五个维度的监管之策。^③同时，网络技术的发展催生了不断涌现的新市场与新型行为方式，^④智慧监管在监管数据的深度利用、监管资源的科学配置中通过实时应答、远程监测、智能传感等对监管对象的“全景

① 参见周济南、苏厚任：《政府算法决策风险的生成机理与防范策略——基于符号互动论的分析框架》，《理论月刊》2025年第1期，第46~55页。

② 参见魏钰明：《大模型应用拓展下的智能社会复杂性及其治理》，《电子政务》2025年第3期，第2~6页。

③ 参见王岭：《数字经济时代中国政府监管转型研究》，《管理世界》2024年第3期，第126~204页。

④ 参见蓝志勇、吴件：《电商时代的协同监管理论之探》，《中国行政管理》2021年第6期，第37~43页。

敞视空间”全面升级了监管效能，提升了监管的精细化、科学化、均衡化程度。^①在新技术的助力下，监管模式已颠覆了传统监管机制的理念与策略，实现从静态到动态、从分散到协同的智慧监管。

技术演化至今，以生成式人工智能为代表的人工智能技术或是人类社会生产力爆发的节点，它既关乎未来福祉，也在各国日益激烈的技术竞争中凸显独特价值。生成式人工智能价值链具有多主体参与和机器自主学习的特征。^②在智慧监管视域下，我国、美国、欧盟等国家和国际组织均形成了各具特色的监管路径。

四、生成式人工智能监管的大国差异

监管是现代国家政府的基本职能和重要职能，它作为优化资源配置和稳定市场秩序的基本制度而不断完善与发展。不同国家监管治理传统、组织结构、决策方式等的差异导致监管的侧重不同。纵观全球，随着生成式人工智能技术的突破发展与相关行业的快速壮大，各国和国际组织已经开始探索富有特色的智慧监管之路：2023年7月，我国率先在国家层面出台《办法》；2023年10月，美国总统拜登签署《发展安全、可靠、值得信赖的人工智能的行政命令》；2023年12月，欧盟就人工智能相关法案达成一致；2024年5月，欧盟理事会正式批准《法案》。2024年成为全球生成式人工智能的监管元年（见表1）。

表1 全球代表性国家与国际组织的生成式人工智能监管特点

	中国	欧盟	美国	英国
特点	提出了对人工智能系统制定详细和具体的规则、标准和要求	专注对人工智能系统可能造成的潜在危害进行分类和优先级排序	规定了人工智能系统的基本原则或指导方针，将实施的解释和具体细节留给组织	专注实现可衡量的与人工智能相关的结果，而不规定必须遵循的具体流程或行动

① 参见叶岚、王有强：《基层智慧监管的政策过程与创新机制——以东部沿海城市区级市场监管部门为例》，《中国行政管理》2019年第8期，第35~40页。

② 参见谢尧雯：《生成式人工智能价值链行政监管与侵权责任的匹配》，《政法论坛》2025年第2期，第36~46页。

续表

	中国	欧盟	美国	英国
优势	降低监管复杂性； 具备执行一致性	针对应用领域而专门定制，与风险水平成比例，具备适应风险水平变化的灵活性	旨在促进发展； 具备监管可塑性； 鼓励和激励创新	鼓励创新发展； 具备一定灵活性； 具备经济效益
挑战	监管刚性过强； 规则难以执行； 可能阻碍市场发展	风险评估较为复杂； 风险情况不停变化； 可能阻碍市场发展	原则与实践具备不一致性，原则难以被应用和执行，原则执行情况难以评估，可能导致行业及市场失控	各主体责任过于分散，结果评估较难执行，可能导致市场秩序失控

（一）我国生成式人工智能的监管策略及特点

生成式人工智能对发展新质生产力具有重要作用，但它具备风险种类多样、损害方式隐蔽、危害结果严重等特点。^① 面对生成式人工智能的系列风险与高速增长的市场空间，我国对相关技术的政策响应与监管治理走在国际前列。2023年4月11日，我国在生成式人工智能领域率先开展工作，就《生成式人工智能服务管理办法（征求意见稿）》公开征求社会意见。业界、学界等多元主体就生成式人工智能的风险及监管展开广泛探讨并提出相应监管治理思路与办法。2023年7月13日，国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局正式颁发以坚持安全和发展并重、促进创新和依法治理相结合为原则的《办法》，以信息内容安全为规范重点，^② 指出将以一揽子有效措施鼓励我国生成式人工智能创新发展，并对生成式人工智能服务实行包容审慎和分类分级监管。

《办法》从技术发展与治理、服务规范、监督检查和法律责任三个重要方

① 参见高志宏：《回应与超越：生成式人工智能法律规制——以〈生成式人工智能服务管理暂行办法〉为视角》，《社会科学辑刊》2024年第5期，第121~130页。

② 参见张学府：《生成式人工智能服务信息内容安全的三类标准——基于〈生成式人工智能服务管理暂行办法〉的分析》，《中国行政管理》2024年第4期，第120~128页。

面对生成式人工智能发展做出规定，既鼓励和支持生成式人工智能技术在各领域和各行业的创新应用，推动构建健康向上的应用生态体系，又从训练数据处理活动、数据标注、开发者责任等角度提出敦促生成式人工智能安全发展的监督原则。^① 梳理共计24条的《办法》尤其是第4章的“监督检查和法律责任”的相关条款可以发现，中国当前对生成式人工智能技术及服务的监管办法具备以下五个方面的特点。

一是联动式系统性监管。《办法》和《中华人民共和国网络安全法》（以下简称《网络安全法》）、《中华人民共和国数据安全法》（以下简称《数据安全法》）、《中华人民共和国个人信息保护法》（以下简称《个人信息保护法》）、《中华人民共和国科学技术进步法》（以下简称《科学技术进步法》）和《互联网信息服务算法推荐管理规定》（以下简称《规定》）等已颁布的法律法规完成对生成式人工智能的兜底式监管，从数据安全、算法推荐、技术服务等角度对生成式人工智能技术、内容、服务涉及的多个领域进行依法监管。

二是基于风险评估的分类监管。《办法》明确指出国家有关主管部门应根据生成式人工智能技术特点及其服务应用，制定科学的分类分级监管规则或监管指引。在国家安全、社会动员、用户隐私、知识产权等方面有不同程度的规定和指导。基于风险的监管以风险为工具，依据不同风险而实施有效干预，确保监管的理性化和科学化。

三是事先设定“风险容量”（risk appetite）。《办法》对涉及国家安全、商业机密、境内外服务等重要领域做出规定，但对生成式人工智能技术及服务的海量应用并未做出明确要求，旨在依据风险级别而对监管对象实施不同强度的监管干预。^② 由于国家不可能防控所有风险，故而在明确可接受的风险（acceptable risks）的基础上，对某些特定风险类别和水平赋予一定容忍性（tolerability），以降低监管成本、优化监管方式、精确实施干预。

四是在由政府主导的多行政部门合作共管的基础上，倡导多元主体协同监管。七部门联合出台的《办法》指出，网信、科技、工业和信息化等部门依据各自职责依法加强生成式人工智能服务管理。同时，《办法》指出，开发者、

① 参见袁佳音：《生成式人工智能服务提供者的侵权责任边界研究》，《中国科技论坛》2024年第9期，第122~133页。

② 参见曾聪：《生成式人工智能与“不可解释性”危害结果的客观归属》，《华中科技大学学报》（社会科学版）2024年第5期，第53~62页。

行业组织、企业、使用者等有关主体在生成式人工智能风险防范等方面展开合作。除了第3章明确提出的提供者服务规范,《办法》同时指出,使用者在发现生成式人工智能服务不符合法律法规时,有权向相关主管部门投诉、举报。

五是突出技术提供者责任。以技术开发者为核心的生成式人工智能提供者应主动承担生产者责任,^①完成自我监管。例如,开发者提供具有社会动员能力或舆论属性的生成式人工智能服务的,应主动按照国家有关规定进行安全评估并依法完成备案等手续。技术提供者应建立健全投诉举报机制,在发现使用者利用生成式人工智能服务从事违法活动时,应采取警示、暂停服务等措施,并主动向有关部门报告。

(二) 欧盟生成式人工智能的监管策略及特点

2021年4月,欧盟在风险评估的基础上率先就人工智能领域展开立法工作,并于2024年3月以压倒性多数通过了全球首份全面系统的人工智能法律框架,即《法案》。《法案》包括人工智能禁用行为、高风险人工智能系统、人工智能系统的透明度义务、通用人工智能模型、创新支持措施、欧盟高风险人工智能系统数据库、市场监督及信息共享、行为准则、授权和委员会程序、处罚等内容,在监管主体、实施路径、管辖标准等方面做出明确规定。^②《法案》根据不同人工智能技术的应用特点将其划分为4种类型:最小风险、有限风险、高风险、不可接受的风险(见表2)。^③其中,生成式人工智能被划分为高风险。《法案》既对以生成式人工智能为代表的高风险人工智能系统分类规则做出严格细致的定义与归类,又明确了对此类高风险人工智能系统的开发、设计、应用和发展相关要求,对高风险人工智能系统的提供者和部署者及其他各方的义务做出系列规定,并建构了以数据、算法等为核心的风险管理系统。^④《法案》具备三个典型特征。

① 参见姚志伟:《生成式人工智能服务提供者在私法上的法律性质》,《上海交通大学学报》(哲学社会科学版)2024年第12期,第127~142页。

② 参见徐伟、张丽梅:《人工智能立法:欧盟经验与中国路径》,《德国研究》2024年第6期,第76~93页。

③ 参见袁曾:《算法应当被解释吗?——人工智能“可控制”的治理向度》,《法学论坛》2025年第1期,第130~142页。

④ 参见曾雄、梁正、张辉:《欧盟人工智能的规制路径及其对我国的启示——以〈人工智能法案〉为分析对象》,《电子政务》2022年第9期,第63~72页。

表 2 欧盟《法案》划分的技术风险等级

风险类型	应用特点	代表产品	监管措施
最小风险	具有一定人工智能技术应用的电子游戏等	手机客户端单一功能电子游戏	不做干预
有限风险	使用者能意识到与机器互动且意志不受干扰的人工智能产品	网页端商品自动回复机器人	透明公开
高风险	可能运用于民主程序及司法系统的人工智能应用，可能用于考试或职业培训、可能决定某人受教育机会的应用等	ChatGPT	开展风险评估，上市前受到严格管控，所有活动记录确保可追溯，上市后持续的市场监督，严格执法和处罚等
不可接受的风险	威胁人类安全、生存及权利，操纵人类行为并规避用户自由意志的人工智能系统等	语音辅助玩具	禁止；若违反规定，则处以上一财年全球最高营业额 6% 的罚款

一是适用广泛的监管范围。《法案》遵循智慧监管倡导的平衡安全与发展的监管理念。一方面，《法案》旨在打造合法、可信、安全的人工智能市场。^①《法案》将为符合欧盟价值观及《欧盟基本权利宪章》规定的基本权利的人工智能系统的开发、投放、服务及使用制定统一的法律框架，促进以人为本和值得信赖的人工智能的应用，防止人工智能系统在欧盟境内产生有害影响，并在此基础上鼓励创新。另一方面，《法案》试图使欧盟以先发优势成为全球人工智能标准的“守门员”。早在 2016 年，欧盟通过的《通用数据保护条例》（General Data Protection Regulation, GDPR）在数据领域的监管及治理体系中取得巨大成效，该条例成为欧盟乃至全球各国几乎无法避开的通行标准。《法案》或将复刻 GDPR 的成功，在欧盟整体技术并不领先全球背景下，通过抢占技术监管治理高地而成为全球人工智能发展的重要参与者与规则制定者。同时，在监管对象上，《法案》明确 7 类适用主体，具备极其广泛的域外适用效力，覆盖了人工智能全产业链及任何与欧盟市场有交集的人工智能系统的

^① 参见陈少威、杨涛、贾开：《比较政策研究视野下全球人工智能治理模式的差异、共识与改革启示》，《中国行政管理》2024 年第 12 期，第 15~24 页。

多元主体。^①

二是以专门办公室为特点的监管主体。值得注意的是，在智慧监管倡导的多元主体协同监管的基础上，《法案》的创新之处在于成立专门办公室，就技术的发展、部署和监管发挥作用。2024年5月29日，欧盟委员会正式成立了由监管与合规部、人工智能安全部、卓越人工智能和机器人部、人工智能造福社会部、人工智能创新和政策协调部等几个专门部门组成的人工智能办公室（AI Office）。人工智能办公室聘用由技术专家、政策专家等构成的专业团队在《法案》运行中发挥关键作用，特别是在通用人工智能大模型方面进行全方位的监督管理。同时，该办公室将促进与欧洲人工智能委员会及成员国的交流合作，支持技术创新并鼓励相关投资。

三是以罚款为主要手段的监管工具。《法案》根据风险等级设计了金字塔型的技术类型划分以及与之相匹配的监管措施，在技术开发、应用和服务的全过程中进行不同等级的监督管理工作。^② 尽管《法案》缜密且详细地设计了诸多条款，然而在监督执行层面仍以罚款为主要惩罚手段，相关条款已引发部分国际科技巨头公司的普遍担忧。

（三）美国对生成式人工智能的监管策略及特点

美国对生成式人工智能的监管特色在于其较为柔和、弹性、宽松的原则导向型监管策略。一直以来，美国在人工智能领域高度强调相关技术作为未来国家发展战略的重要性，并未在国家层面颁布以管控为目标的人工智能法案，旨在将美国打造为全球人工智能发展的领导者。美国的监管策略具备以下特点。

一是监管体系的分散性、保守性。美国就生成式人工智能领域的监管策略可谓“雷声大、雨点小”。美国州政府层面分别就隐私保护、医疗保健、算法歧视、数据安全等方面提出相关立法建议和监管要求，并在以生成式人工智能技术为代表的人工智能等领域展开局部性监管措施的探讨与尝试。在联邦政府层面，美国就国家安全、公民权利、创新发展等领域提出综合性整体

① 参见吴艳、刘腾龙：《欧盟人工智能安全治理的路径与前景》，《湖北社会科学》2024年第12期，第32~41页。

② 参见吴建南、马太平、周磊：《人工智能安全治理的规则体系：国际比较与协同进阶》，《中国行政管理》2024年第12期，第6~14页。

框架，推出了《创造实验人工智能资源法案》等。从立法、司法、执法三个层面看，虽然美国两院和联邦政府等就生成式人工智能相关法案发起层出不穷的讨论，但是真正落地和执行的法案及监管措施少之又少，尚未有类似欧盟《法案》这样全域性、全面性的法案。

相较我国和欧盟，美国尚未出台全域性、全面性的基于风险进行严格分类与缜密监管的强硬法案，只是针对人工智能技术设计的部分领域进行单点式立法调整，以“打补丁”的方式延续原有的技术监管路线，^①保持了自2016年奥巴马政府以来在人工智能领域以鼓励创新为基本原则的政策的一致性。虽然围绕公民信息安全、数据保护等领域美国已举办若干场影响广泛的听证会，但是在热切讨论之外并没有进一步的国家层面的立法动作或执行动作，传统监管思维及监管工具是否能够匹配与适应生成式人工智能及相关人工智能技术仍旧存疑。2023年5月，在生成式人工智能引发全球震动后，OpenAI首席执行官萨姆·奥尔特曼（Sam Altman）在出席美国国会听证会时指出，先进技术引发的虚假信息、有害内容、劳动者保护等监管议题已是数字时代以来的老生常谈。生成式人工智能的出现既不意味着原有监管工具的必然失效，也不见得需要政府大幅度超越原有人工智能技术的监管策略和监督模式。

二是鼓励发展、审慎干涉的监管特点。在强调市场主导、创新驱动、自由竞争、安全可靠的监督原则基础上，美国对生成式人工智能领域的监督管理具备高度的灵活性、开放性、适应性。既避免对潜在技术风险的不安全感过度反应与过度处理，^②又避免从国家层面以法律等强制性手段干扰甚至打破生成式人工智能技术与行业的创新性发展。

美国注重政府、市场、社会层面的沟通与协作，^③在对产业实践的充分论证和调研了解的基础上缓慢且审慎地陆续出台相应的监管政策。一方面，大型科技公司深度参与了美国对技术监管的政策制定，在某些程度上确保了以发展为首的监管目标。另一方面，从更为宏大广阔的时空环境看，美国独特

① 参见贾夏利、刘小平：《中美人工智能竞争现状对比分析及启示》，《世界科技研究与发展》2022年第4期，第531~542页。

② 参见陶于祥、吴超楠、李晶莹等：《基于技术生命周期的中美人工智能原始创新能力研究》，《中国电子科学研究院学报》2021年第12期，第1215~1223页。

③ 参见刘玉书：《中美人工智能战略及政策的比较研究》，《云南行政学院学报》2022年第1期，第101~124页。

的文化背景、政府结构、政策出台过程以及市场格局也形塑了其在智慧监管视域下重发展、轻工具、缓节奏的监管格局。

（四）英国生成式人工智能的监管策略及特点

英国致力于通过合理有效的智慧监管方式实现有利于生成式人工智能等先进技术发展的优良环境。从《国家人工智能战略》《建立有利于创新的方式来监管人工智能》等报告到2023年正式发布的《白皮书》，英国反对不合理的过度监管，避免设置复杂的监督机制，试图通过现有法律制度发挥有效的适度的监管作用。相较欧洲部分国家的技术恐慌，英国政府指出其构建的监管体系不以追求应对广泛的社会风险与全球性挑战为目标，而是要鼓励英国本土人工智能技术的快速发展，以尽快实现英国在人工智能领域的全球领导者地位。具体的监管特点如下。

一是最大限度避免监管干扰。英国向来重视科技伦理的合理有效监督，^①《白皮书》建构了由政府、企业、第三方监督机构多元主体协同监管的策略，通过发挥不同主体的优势与特性确保为生成式人工智能等技术的发展提供保障。值得注意的是，相较欧盟《法案》，《白皮书》明确提出将不会就人工智能创建新的监管机构。在承认政府在监管系统中存在必要性的基础上，英国将最大限度地保护人工智能的快速发展，政府将在《监督机构法典》等现有法律法规基础上建立更为灵活友善、高效运转的监管体系，将不会额外就生成式人工智能等技术展开新的大幅度的监管布局。一方面，英国将明确不同主体对应的监管边界，避免职权交叉或某些监管主体的权力扩张。另一方面，《白皮书》指出，应避免监管机构对人工智能领域的反复介入、重复要求与过度干扰。

在监管范围和监管方式上，《白皮书》明确列出确保监管的适当性与科学性的相关条款。一方面，对以生成式人工智能为代表的技术在特定场景下产生的某些对应性结果进行精准监管，《白皮书》强调需要保证支持技术创新的监管架构的合理性与灵活性，避免技术迭代发展而导致的政策僵化与过时。另一方面，相关监管机构只能在其法定专业范围内进行小范围小幅度的监管及执行活动，其监管模式须遵循五个基本原则，即安全性和稳定性、透明性

^① 参见宋应登、霍竹、邓益志：《英国科技伦理监管体系研究与启示》，《中国科技论坛》2024年第8期，第179~188页。

及解释性、公平性、问责制度与治理机制、竞争和补救措施。

二是最大限度鼓励技术发展。英国政府在监管理念方面强调采用灵活的监管模式，以促进政府、市场和社会之间的合作，并确立了六个监管特征。第一，支持创新，推动技术蓬勃发展。第二，注重相称性，避免监管措施对企业和相关机构造成不必要或不成比例的压力和负担。第三，强调可信赖性，在以结果为导向的监管模式下解决实际风险，并培养社会公众对人工智能技术的信任与支持。第四，着眼适应性，各方主体，特别是英国政府要迅速适应人工智能技术发展带来的机遇与挑战。第五，明确规定适用于人工智能生命周期所有参与者的监管规则。第六，强调协作性，鼓励政府和多元主体共同参与监管，建立信任和可靠的人工智能发展生态系统。

全球主要大国与国际组织已在人工智能监管与治理实践中形成各具特色的模式与体系。在智慧监管视野下比较我国、欧盟、美国、英国就生成式人工智能的监管方案及监管特点，可以发现各国在监管主体的组成上都包括由政府、市场等组成的多元主体，只有欧盟单独成立全新的人工智能办公室，而英国等基本保持了原有监管主体。各国及国际组织在监管工具的使用上虽有类型上的分化与调整，但基本遵循了不同层级的由硬性法律法规到软性警戒措施的工具组合。可以进一步将各国的监管之道划分为两种不同目标导向的监管策略：一是以我国、美国、英国为代表的鼓励发展型监管；二是以欧盟为代表的管制风险型监管。总体来看，技术演化至今，以生成式人工智能为代表的新兴技术或许是人类社会生产力爆发的关键节点，它既关乎未来福祉，也在各国日益激烈的技术竞争中凸显独特价值。全球对生成式人工智能的监管仍在探索和发展之中，需要在当前基础上持续更新方法与策略以保障生成式人工智能技术的向善发展。

五、未来生成式人工智能智慧监管实践的发展进路

在生成式人工智能的影响和冲击下，对其监管策略及监管过程在政府及非政府主体的不断互动中存在适应性调整的过程，处于全新阶段的新兴技术

构成了国家数字经济崛起的核心驱动力，^①然而人工智能的法律治理仍处于规则形成期。^②既没有一劳永逸的监管方案，也不存在适配所有情况的监管办法。

（一）当前生成式人工智能监管实践的不足

在国际视野下，当前对生成式人工智能技术及服务的智慧监管仍存在五个方面的不足，需要加以应对与突破。

一是分类分级监管模式的回应性缺陷。因生成式人工智能的不同风险等级而有针对性设计的监管模型难以应对在复杂情境中动态变化且深度交互的监管现实，监管原则、监管主体、监管对象、监管工具、监管制度之间存在张力。一方面，风险等级的分级分类存在较多挑战和争议。生成式人工智能技术已逐步深入各个领域，每一个子领域都涉及海量指导性细则。监管网络过密则影响技术发展，过疏则引发风险逃逸。同时，即便相同风险等级的监管对象也存在差异极大的发展特性与运作风格。^③另一方面，低风险与高风险之间存在模糊地带和动态转化。生成式人工智能技术的快速迭代和智能应用会使不同服务的风险等级处于持续性的变化和调整中，分类分级的调试速度或难以匹配现实变化。

二是监管俘获与技术黑箱。以欧盟的《人工智能法案》为例，它对生成式人工智能的技术开发者、服务提供者乃至用户都提出了相应的监管责任与义务划分方案。一方面，基于密切关系的自由裁量权机制显著增加了监管俘获风险，尤其在科技巨头话语权不断扩张的背景下，被监管方可能通过资源渗透将监管主体纳入利益网络。监管边界的模糊性使科技企业得以游走于政策弹性区间，甚至反向塑造部分监管规则。另一方面，过度信任非政府监管主体或为监管带来不符合监管最终目标及多元主体利益的后果。算法不仅具备工具层面的赋能作用，而且呈现并建构着社会结构中的权

① 参见赵精武：《人工智能治理中的技术自主权》，《中国社会科学院大学学报》2025年第1期，第94~112页。

② 参见张凌寒：《人工智能法律治理的路径拓展》，《中国社会科学》2025年第1期，第91~110页。

③ 参见 Henry Rothstein, Olivier Borraz and Michael Huber, Risk and the Limits of Governance: Exploring Varied Patterns of Risk-Based Governance Across Europe, *Regulation and Governance*, Vol. 7(2), 2013, pp. 215-235。

力关系。^① 在技术壁垒极高的生成式人工智能领域，服务提供者具备天然的话语优势与主导权，它的研发和应用的整个过程处于高技术门槛的“黑箱”之中。

三是监管漂移与监管逃逸。生成式人工智能技术、内容、服务、行业充满复合且复杂的监管交叉地带，对某一议题的界定和判断往往需要多部门联合给出针对性意见。以我国的《办法》为例，它并未划定各部门的监管职责与权限，不同监管主体权责难以划分，既容易偏离相应法律法规设定的监管边界，又容易使监管机构在容易解决的简单问题上一拥而上，而在较有争议且难以处理的问题上踌躇不前。这种两极化的监管行为会导致生成式人工智能在某些方面被过度监管，而在某些范畴逃逸现有监管结构。生成式人工智能以其灵敏性、智能性、弥散性、迭代性而不同于既往人工智能技术及新兴产业，一进一退、一张一弛之间极易导致监管失效。

四是自我参照视域下的结构耦合。以欧盟的《法案》及英国的《白皮书》为例，其监管办法均未明确规划不同系统的合作机制。一方面，以经济和法律等为代表的子系统具备自我的对话语言和系统理性，既会对系统外的因素做出反应与反馈，又在日渐封闭的系统话语里难以与其他系统兼容或和谐合作；另一方面，不同的行政部门，如安全部门、宣传部门等既存在多部门间横向上的错位，又存在部门内部纵向上的差距，在不同部门利益、部门规范、规范风格的情况下，这种条块分割加剧了跨部门合作的难度，不同系统的结构耦合或许会产生监管失灵的困境。

五是监管干预导致的预期外影响。监管制度中监管者与被监管者之间乃至监管者内部与被监管者内部存在张力，这种主体的对抗性和监管干预的局限性会加剧监管结果的偏移，可能引发一系列与监管设计预期结果相违背的结果。一方面，监管活动或对生成式人工智能技术和产业发展造成干扰，尤其是当前我国乃至全球对可行或不可行的界定仍较模糊，这会极大影响生成式人工智能的创新发展。另一方面，不同监管机构在权力扩张、绝对安全等目的下会刻意放大或过度服从刻板规则，导致监管的工具价值取代促进生成式人工智能健康发展的最终价值，这种目标置换或使得生成式人工智能相关领域成为监管主体扩张的工具和砝码。同时，监管中对不同风险等级的标签

^① 参见臧雷振、陈浩：《生成式人工智能算法风险及社会治理挑战》，《中共中央党校（国家行政学院）学报》2025年第1期，第43~53页。

化分类或对原本旨在规避的风险类型形成指引和参照，技术提供者或可刻意避免关键词、关键步骤而成功绕过已经成型的监管框架，反而造成了负面效应的隐秘放大。

（二）我国生成式人工智能智慧监管的可能进路

在习近平新时代中国特色社会主义思想的指导下，通过全面推进市场监管现代化，我国已初步构建统一、有序、高效、安全、和谐的超大规模市场，并不断加强多元主体协同监管，市场秩序持续好转。^①党的十八大以来，人工智能产业作为新质生产力的核心要素得到了充分的重视和支持，我国先后出台了多份中央层面的文件以引导、支持、保障人工智能领域的高质量发展。在国际比较视野下，我国未来对生成式人工智能的智慧监管可以借鉴欧盟、美国、英国的监管特点，以保障和促进生成式人工智能的高质量发展。

一是要丰富监管主体，形成监管合力。在监管主体的多样性与专业性上，可借鉴欧盟。欧盟的人工智能办公室承担了以生成式人工智能为核心的人工智能技术及服务安全发展的重要监管工作。在监管主体多元化的基础上，成立专有机构有助于抽调精准力量、聚集专业人才、统筹监管方案，在避免条块分割“踢皮球”“打太极”的同时，从理念、模式、机制三个方面优化监管路径，^②形成长期的、稳定的、迭代的监管格局。

2024年1月，工业和信息化部、教育部、科技部、交通运输部、文化和旅游部、国务院国资委、中国科学院联合印发《关于推动未来产业创新发展的实施意见》（以下简称《意见》）。《意见》指出：“未来产业由前沿技术驱动，当前处于孕育萌发阶段或产业化初期，是具有显著战略性、引领性、颠覆性和不确定性的前瞻性新兴产业。大力发展未来产业，是引领科技进步、带动产业升级、培育新质生产力的战略选择。”^③大模型、超大规模新型智算中心等前瞻性赛道和标志性产品是未来产业的核心所在。《意见》提出以开放合

① 参见刘鹏、郭戈英：《走向适应性监管：理解中国金融监管机构改革的治理逻辑》，《学海》2023年第3期，第42~50页；刘亚平、苏娇妮：《中国市场监管改革70年的变迁经验与演进逻辑》，《中国行政管理》2019年第5期，第15~21页。

② 参见郭江兰：《高风险人工智能监管的批判与超越》，《北京工业大学学报》（社会科学版）2025年第3期，第48~60页。

③ 《工业和信息化部等七部门关于推动未来产业创新发展的实施意见》，https://www.gov.cn/zhengce/zhengceku/202401/content_6929021.htm，2024年12月1日。

作、安全有序的基本原则营造包容审慎的发展环境，推进全面布局未来产业、加快技术创新和产业化、打造标志性产品、壮大产业主体、丰富应用场景、优化产业支撑体系等六大重点任务，探索跨部门联合治理模式，建构多元参与、有效协同的未来产业治理格局。^①2024年5月，工信部高新技术司设置了未来产业处，这既是国家部委首次以未来产业为核心而专门设置的业务主管机构，也是就新兴技术与治理方案的全新尝试与探索。大国具备资源支配能力、标准制定能力与话语构建能力，^②可以参考欧盟的人工智能办公室及我国新成立的未来产业处，就生成式人工智能等新兴技术成立专有机构，有助于形成部际协同、央地协作的工作格局。

二是要融汇监管基础，形成监管网络。在监管网络和监管空间上可参考英国。人工智能对未来地缘政治形态的塑造具有关键作用，将成为大国竞争的关键要素。^③英国政府指出，将结合既往监管基础及格局，对生成式人工智能等新技术新产业采取更为灵活的监管策略，旨在安全有序地保障和促进技术发展。循迹过往，我国已具备较为成熟、全面和成体系的人工智能技术与治理脉络，可在此基础上循序渐进地精细化推动生成式人工智能的监管进程。

2016年至2024年，我国已就人工智能的理论研究、技术创新、产业发展、政策配套等维度展开全面布局。2016年颁布的《中华人民共和国国民经济和社会发展第十三个五年规划纲要》已正式提及“人工智能”这一概念，并强调深化科技创新引领作用。^④2018年10月31日，习近平总书记在主持中共十九届中央政治局第九次集体学习时强调，“人工智能是引领这一轮科技革命和产业变革的战略性技术”，“加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手”。^⑤2020年，《中共中央关于制定国民经济和社

① 参见《工业和信息化部等七部门关于推动未来产业创新发展的实施意见》，https://www.gov.cn/zhengce/zhengceku/202401/content_6929021.htm，2024年12月1日。

② 参见宫云牧：《技术权力视角下的中美欧人工智能技术竞争》，《欧洲研究》2025年第1期，第24~52页。

③ 参见汪道、李国强：《美国智库对人工智能影响国际关系的观点与启示》，《智库理论与实践》2025年第1期，第19~28页。

④ 参见《中华人民共和国国民经济和社会发展第十三个五年规划纲要》，人民出版社2016年版，第17~77页。

⑤ 《习近平在中共中央政治局第九次集体学习时强调 加强领导做好规划明确任务夯实基础 推动我国新一代人工智能健康发展》，《人民日报》2018年11月1日。

会发展第十四个五年规划和二〇三五年远景目标的建议》指出：“把科技自立自强作为国家发展的战略支撑，面向世界科技前沿、面向经济主战场、面向国家重大需求、面向人民生命健康，深入实施科教兴国战略、人才强国战略、创新驱动发展战略，完善国家创新体系，加快建设科技强国。”^①《国际新一代人工智能标准体系建设指南》《国家新一代人工智能创新发展试验区建设工作指引》等文件进一步指出在“统筹规划，分类施策，市场驱动，急用先行，跨界融合，协同推进，自主创新，开放合作”原则下建立国家新一代人工智能标准体系，以促进中国人工智能技术及产业蓬勃发展。2021年，国务院印发《“十四五”数字经济发展规划》，指出我国需要“坚持公平竞争、安全有序。突出竞争政策基础地位，坚持促进发展和监督规范并重，健全完善协同监管规则制度”。^②2022年，我国陆续发布《规定》《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》《工业和信息化领域数据安全管理办法（试行）》等文件，就人工智能领域算法、数据、场景三项核心要素展开监督管理工作。至此，我国就人工智能领域的基本要素、基本应用、基本场景等均出台了局部性或全面性的监督与管理办法，为生成式人工智能这一人工智能领域的最新突破筑牢了监督基础，提供了监督依据，铺陈了监督网络。

2023年正式发布的《办法》既运行于现有政策的基础上，又与《网络安全法》《数据安全法》《个人信息保护法》《科学技术进步法》等法律法规构成了协同性体系性监管制度。生成式人工智能仍在高速发展中，但2022年底至2024年6月，我国乃至全球并未有以此技术为基底而出现的大型风险，仍在人工智能技术甚至新兴技术的固有风险阈值内。因此，可借鉴英国形成灵活、弹性，甚至“缓慢”的监管政策，在给予技术发展空间的同时以修补的方式筑牢监管之基。

三是要创新监管理念，协调监管机制。在监管理念方面，可借鉴美国。2023年10月，美国总统拜登签署涵盖人工智能发展与安全等多个维度的行政命令，强调以生成式人工智能为代表的人工智能技术的重要性，旨在将美国

① 《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》，人民出版社2020年版，第10页。

② 《国务院关于印发“十四五”数字经济发展规划的通知》，《中华人民共和国国务院公报》2022年第3期，第7页。

打造为全球人工智能领导者。在此理念引导下，美国对人工智能采取较为柔和的分布式、分散式监管，在保证创新发展的原则下审慎处理生成式人工智能可能带来的系列挑战与风险讨论多而行动少、看法多而落实少等问题，为生成式人工智能的发展提供了一定阶段和一定程度的窗口期与空白期。

尽管当前对技术进行风险等级划分并匹配相应监管工具是较为通行的做法，然而基于风险的监管方式在便于操作执行的同时，前文提及的回应性缺陷、技术黑箱等问题也普遍存在于各国的监管模型中。第一，分类存在交叉。例如，根据技术要件、应用场景、社会影响等不同侧重，同一技术产品或被划分子不同风险等级。过细过密的分类很可能华而不实，风险分类有效性低甚至互相冲突。第二，技术合规成本骤增，在冗杂的风险条款中难以找到生存空间，不利于技术创新发展。结合2016年至今关于人工智能及创新发展的相关文件，在当前国内外发展态势的背景下，需要坚持系统思维、坚持创新引领、坚持安全发展、坚持市场主体地位、坚持包容审慎监管、坚持多元共治的监管理念，以驱动生成式人工智能的技术向善与健康发展。

整体上，生成式人工智能技术仍在快速迭代优化，生成式人工智能市场正在剧烈增长和发展，技术革命性突破、生产要素创新性配置、产业深度转型升级而催生的以生成式人工智能为代表的生产力，是高质量发展的内在要求和重要着力点。习近平总书记深刻指出，“我们必须牢记高质量发展是新时代的硬道理，完整、准确、全面贯彻新发展理念，把加快建设现代化经济体系、推进高水平科技自立自强、加快构建新发展格局、统筹推进深层次改革和高水平开放、统筹高质量发展和高水平安全等战略任务落实到位”，“新质生产力的显著特点是创新，既包括技术和业态模式层面的创新，也包括管理和制度层面的创新。必须继续做好创新这篇大文章，推动新质生产力加快发展”。^①新技术的发展也呼吁新型监管格局的建构。在比较视野下，我国现有的监管体系与模式仍在与生成式人工智能等新兴技术的互动、磨合甚至博弈中，仍需要密切观察技术创新、行业发展、市场机制乃至国际格局，动态、灵活、审慎地调整监管策略。需要把握好对生成式人工智能监管的尺度与效率，确保生成式人工智能的创新与发展。

（责任编辑：方 军）

^① 习近平：《发展新质生产力是推动高质量发展的内在要求和重要着力点》，《求是》2024年第11期，第5、6页。