

生成式人工智能敏捷治理的国际趋势与建构模式

迟舜雨¹, 杜安琪²

(1. 中国政法大学法与经济学研究院, 北京 100088;

2. 中国信息通信研究院, 北京 100191)

摘要: 敏捷治理作为一种新的治理范式, 有助于提升生成式人工智能治理水平。通过分析生成式人工智能的治理需求与发展挑战, 揭示了传统治理模式的不足, 并认为以多元共治、灵活的治理依据和迭代过程为核心特征的敏捷治理能够有效平衡技术创新与风险控制的需求, 可以弥补传统治理范式在生成式人工智能治理的局限。在审视了国际敏捷治理的实践后, 提出了适用于中国的生成式人工智能敏捷治理策略建议。为保障中国生成式人工智能的高质量发展和高水平安全, 应在多元参与的基础上进一步发展多元共治模式, 加强硬法与软法之间的互动, 并逐步丰富生成式人工智能的治理工具箱。

关键词: 生成式人工智能; 敏捷治理; 多元共治; 去规则化; 迭代流程

中图分类号: TP18, D035, F49

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202434

International trends and constructive models of agile governance for generative artificial intelligence

CHI Shunyu¹, DU Anqi²

1. School of Law and Economics, China University of Political Science and Law, Beijing 100088, China

2. China Academy of Information and Communications Technology, Beijing 100191, China

Abstract: Agile governance, as a new paradigm of governance, is conducive to enhancing the governance level of generative artificial intelligence. By analyzing the governance needs and development challenges of generative artificial intelligence, the inadequacies of traditional governance models were revealed. It was believed that agile governance, characterized by multi-stakeholder co-governance, flexible governance basis, and iterative processes, could effectively balance the needs of technological innovation and risk control, and compensate for the limitations of traditional governance paradigms in the governance of generative artificial intelligence. After examining the international practices of agile governance, suggestions for agile governance strategies suitable for generative artificial intelligence in China were proposed. To ensure the high-quality development and high-level safety of generative artificial intelligence in China, it is necessary to further develop a multi-stakeholder co-governance model on the basis of multi-participation, strengthen the interaction between hard law and soft law, and gradually enrich the governance toolkit of generative artificial intelligence.

Key words: generative artificial intelligence, agile governance, co-governance, deregulation, iterative process

0 引言

作为新质生产力的重要驱动力, 生成式人工智

能 (artificial intelligence, AI) 不仅为数字经济的高质量发展注入了新活力, 也带来了复杂的伦理、法律和安全问题^[1]。生成式 AI 是 AI 技术的分支,

收稿日期: 2024-04-18; 修回日期: 2024-07-29

通信作者: 迟舜雨, shunyu.chi@cupl.edu.cn

基金项目: 国家社会科学基金项目 (No.22CFX055)

Foundation Item: The National Social Science Foundation of China (No.22CFX055)

可以用于文本、图片、音频、视频等多个领域内容的生成。近年来，以 ChatGPT 为代表的生成式 AI 的崛起引领了人工智能技术发展新趋势，并开始广泛应用于各行各业，展现出较大的应用潜力和经济影响。治理模式事关产业和国民经济兴衰^[2]。传统治理模式在应对技术快速迭代及其造成的不确定性风险时存在局限，不仅限制了创新的潜力，也可能导致新的风险得不到有效管理。因此，政府需要更灵活的治理范式来引领政策的制定与监管^[3]，而敏捷治理提供理想治理范式^[4]。

敏捷治理的概念最初源于软件开发领域，强调相较于传统的按照既定计划、流程和文档的模式，软件开发应当遵循互动、合作的原则并积极响应变化。敏捷治理被系统引入政府治理始于 Schwab 于 2018 年发表的文章^[5]，随后被世界经济论坛白皮书所借鉴^[6]。敏捷治理的提出以第四次工业革命为背景，一方面，AI 等新兴技术为经济发展带来前所未有的潜力和机遇，另一方面这些技术的应用和破坏式创新（disruptive innovation）^[7]，在数据、算法以及内容的生成阶段均伴随高度不确定的风险^[8]。严格的治理有助于防范风险，但也会抑制创新；而过于宽松的治理环境一定程度上会激发创新潜能，但会损害社会福祉。因此，敏捷治理的核心目标是在促进科技创新与应对风险间寻求平衡。通过构建敏捷治理范式，政府能够在 AI 技术和商业模式加速迭代的背景下更好地统筹发展与安全，推动 AI 高质量发展。

敏捷治理范式并非固定的规则或模式，而是一种灵活、弹性的治理思维，并涵盖一系列治理方法和工具。本文基于传统治理范式所面临的挑战，探

索适用于生成式 AI 的敏捷治理的治理范式、主要特征和国际实践，旨在提出我国生成式 AI 治理的具体建议，为我国在该领域的治理提供参考和借鉴。敏捷治理的发展与建构如图 1 所示。

1 敏捷治理与多元共治

生成式 AI 的巨大发展潜能与潜在风险呼唤社会各界共同参与其治理。多元共治是敏捷治理范式的主要特征。相较于传统较为单一的治理主体，敏捷治理范式下的多元共治模式吸纳包括政府、行业协会、企业、学术界以及公众等多方的广泛参与和共同治理，更好地回应了新时代下生成式 AI 的治理需要。基于此，本节将重点分析多元共治相较于传统模式的优势，并结合国际实践探讨其建构模式。

1.1 从单一主体到多元共治

传统治理范式往往强调政府在社会治理和资源分配中的作用，预设政府与其他主体之间的不平等地位。政府通过制定政策、执行法律、维持秩序等手段，采取自上而下的命令-控制式监管（command and control），以惩戒手段督促其他主体被动服从政府命令。

然而，此范式难以回应生成式 AI 的治理需求。其一，在实现“加快推动人工智能发展”等治理目标方面存在挑战。我国高度重视 AI 发展，2023 年 4 月中共中央政治局召开会议，指出“要重视通用人工智能发展，营造创新生态，重视防范风险”；2023 年 12 月中央经济工作会议强调要“加快推动人工智能发展”；2024 年《政府工作报告》明确指出要“深化大数据、人工智能等研发应用，开展

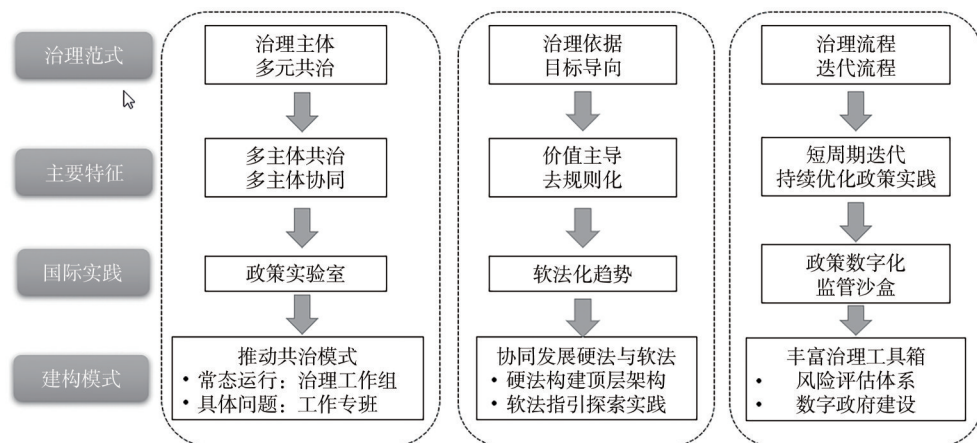


图1 敏捷治理的发展与建构

“人工智能+”行动”。上述目标的实现有赖于政府与各主体之间的良性互动，政府在制定决策和监管时，应与企业进行有效沟通，积极回应生成式AI发展的实际需求，并与企业、行业协会、学者等共同研究和探讨可能存在的风险以及相应解决策略。但在传统治理范式下，监管机构倾向于将监管对象视为需要打击和控制的目标，这种做法不利于形成积极互动的局面，也难以有效达成保障发展、促进创新的政策目标^[9]。其二，在应对AI风险的不确定性方面存在局限^[10]。随着AI技术的迅猛发展，政府与企业均面临难以准确预测和评估技术演进方向及其潜在风险的困境，即所谓的“共同无知”现象。这一现象源于AI技术固有的复杂性、应用场景的广泛多样性以及伴随的风险与法律问题的不确定性。例如，AI系统的决策过程往往呈现出“黑箱”特性^[11]，体现为企业与政府及公众对于技术原理的不了解，因此政府需要通过算法备案等形式提高算法透明度，深入了解AI技术的研发应用，以采取有效的治理措施。然而，生成式AI技术改变了对传统算法黑箱的认知和理解，算法“黑箱”问题的本质由人与人之间的信息不对称转变为人类在AI面前的共同无知。迈向通用AI时代，治理者不仅要关注算法的透明度，还需探索新的治理策略和框架，以应对技术所带来的复杂性和不确定性。

敏捷治理倡导政府、企业和社会各界广泛参与和共同治理^[12]，可以在一定程度增强各方对生成式AI的理解，保障其健康发展。敏捷治理模式不仅涵盖了不同政府部门间的协调治理，还包括政府与企业、研究机构、非政府组织以及公众的合作共治，从两个层面契合了生成式AI的治理需求。其一，多元共治可以确保政策的制定和实施过程中汇集多元的知识、资源和视角。在生成式AI的创新和应用中，需要跨学科、跨领域的知识技能，也面临包括伦理、隐私、安全等一系列风险^[13]。敏捷治理通过推动多方利益相关者的参与，不仅能够促进创新思维的碰撞和技术解决方案的迭代与优化，也有助于应对“共同无知”现象，及时识别和应对潜在风险，实现技术发展和风险管控的平衡^[14]。其二，多元共治有助于增强政策制定和执行的透明度，并提升公众对技术发展方向认同感，打破传统监管者与被监管者之间的典型的“猫和老鼠”博弈关系^[15]。各主体间持续的交流与反馈能够为生成式AI的健康发展创造更加稳定和包容的社会

环境。

1.2 去单一化发展：多元共治的模式与实践

相较于传统单一的治理模式，多元共治更符合AI的治理需求。基于此，本节将具体探讨多元共治的构建模式。

多元共治模式致力于跨越政府部门之间，政府与企业、研究机构以及非政府组织之间的传统边界，采用系统思维（systemic thinking），强调生成式AI治理中的整体性、关联性和动态性。这种思维要求政策制定从宏观层面理解和应对挑战，而非仅着眼于问题的单个方面或表面症状。它鼓励从整体上审视生成式AI带来的挑战和机遇，识别系统内各元素的相互作用和反馈循环以及这些相互作用随时间的动态变化^[16]，进而制定更有效和可持续的治理策略。

因此，多元共治强调各部门、各领域的协同投入，明确各主体在治理中的角色与职责。具体而言，多元共治有两个显著特征。其一，政府、企业、行业协会等作为治理主体，各自承担明确的职责和任务。政府发挥其宏观统筹优势，负责制定宏观政策、监管规则和法律框架，确保治理体系的有序运作。企业在技术创新、市场前沿洞察和自律标准化方面具有优势，应当探索内部治理与合规体系的构建，并积极参与法律法规和政策的制定，在符合技术创新和自身发展需求的同时，推动技术应用的合规、安全与伦理。其他机构，例如行业协会、科研机构等亦应当发挥其关键作用，通过制定行业标准 and 行为准则，组织行业内和学界的培训和交流，提升整体的自律水平和专业能力。其二，在多元共治模式下，各主体之间需要建立常态化交流、沟通与共享机制，确保信息流通，促进协同合作，通过定期的会议、研讨会和联合项目，各主体可以共同研究和解决生成式AI的问题与挑战，提升政策和治理的有效性和适应性。

政策实验室是多元共治的典型操作模式。政策实验室在政府内部聚集具有多样化背景的专业团队，运用创新的方法探索公共政策问题的新方案，旨在助力政府跳出传统治理模式的框架，为政策的制定和执行提供更具创新性和实效性的建议^[17]。英国的政策实验室于2014年发起成立，隶属于内阁办公室（Cabinet Office），由政府官员、社会科学家、设计师、数据分析师以及来自学界和企业的专家组成。这种团队结构鼓励不同背景的专业人士

共享知识，促进各方对新兴技术带来的契机和挑战的全面理解。为了达成多元治理，政策实验室一方面定期组织和举办工作坊、研讨会等，邀请不同政府部门和外部利益相关者参加，为参与者提供了一个共同讨论、理解政策的平台，促进跨部门和跨行业的沟通与合作；另一方面利用数字平台和社交媒体工具促进实验室与外界的互动和合作，通过在线调查、社交媒体互动和虚拟会议等手段广泛收集利益相关者的意见，提升政策制定的透明度。美国的敏捷政府中心（Agile Government Center）发挥类似政策实验室的功能。区别于英国政策实验室为政府机构，美国的敏捷政府中心是由美国国家行政科学院（National Academy of Public Administration）协同包括IBM在内的合作伙伴发起的非营利、非党派的独立组织。虽然敏捷政府中心非政府机构，但与政府部门密切合作。自成立起，敏捷政府中心主张政府部门间应当分享数据、技术和最佳实践，并倡导建立跨部门工作团队，汇集不同领域专家解决新兴技术带来的复杂治理问题。相较于英国政策实验室，敏捷政府中心亦承担培训职能，帮助政府工作人员掌握敏捷合作和沟通的技能。在面临公共问题时，敏捷政府中心除组织研讨会外，也积极倡导公私合作，充分发挥企业等私营部门在社会治理中的主观能动性，利用私营部门的创新能力和资源解决公共问题。

以政策实验室为代表的多元共治模式不同于传统政府治理中的多元参与理念，后者主要强调在治理中参考各方意见，而多元共治模式则更注重各主体的自治、公私合作、跨学科合作，激励各方在治理中的能动性并促进信息、知识的共享。这种模式在生成式AI治理中具有显著优势，可提升政策制定、实施和评估的整体性、关联性和动态性。然而，需要注意的是，政策实验室并非万能，各国仍在探索其适用范围^[18]。由于政策探索和实验具有不确定性，政府在制定政策时通常不愿将紧迫或优先事项分配给政策实验室^[19]。但毫无疑问，政策实验室为多元共治以及敏捷治理模式的构建提供借鉴思路。

2 敏捷治理与灵活的治理依据

传统治理中依赖政府制定和执行规则来实现治理目标，但规则的僵化和滞后性导致其难以应对技术的快速迭代，也可能限制技术的潜在创新。因此，敏捷治理主张去规则化，直接以目标为导向，

采用更灵活和弹性的政策制定和执行方式来实现治理目标。基于此，本节将探讨敏捷治理范式下目标导向的必要性，并结合国际实践探索目标导向的建构模式。

2.1 从规则导向到目标导向

传统治理范式以规则为导向，将事先设定的法律、法规及部门规章作为政府行为和决策的依据，通过政府对规则的遵守和执行实现治理目标。这种基于规则的治理范式不仅可以降低可能的权力滥用风险，也为各方设定了清晰的预期，提供了稳定且明确的指引。因此，世界银行将法律规则的有效性和适应性作为评估良好治理水平的关键标准。

规则导向的治理范式的有效性依赖于政府通过制定和执行规则来实现治理目标。该范式体现出事前规制的特征，通常是对已知风险和挑战的回应，反映了对历史经验的总结和应用。因此，规则导向的治理范式在处理已知问题时具有优势，但面临快速迭代的新兴技术、不断变化的市场环境以及频繁出现的未知风险时，其弹性的缺失往往导致其难以实现治理目标。生成式AI技术的复杂性和应用场景的多样性导致其风险种类多样。对于已知风险，若是技术提供者、服务提供者或者使用者恶意利用技术所造成的风险，如训练数据选择所导致的偏见歧视、使用者利用生成式AI创建高度逼真的钓鱼邮件进行恶意攻击等，则可以在现有法律规范体系内寻找解决方案或者制定新的立法；若是由生成式AI自身技术缺陷带来的风险挑战，如AI“幻觉（hallucination）”是生成式AI面临的一项技术问题，即模型可能生成看似合理但事实上错误或无意义的信息^[20]，这种虚假或误导性信息的输出难以仅靠确定的法律法规解决，还需要探索科技伦理的指导、自律组织的作用等，践行“边干边学”的实验主义理念，通过不断试错去摸索恰当的解决方案。

传统治理模式以规则为导向，强调通过政府在相对固定的体系中制定并执行规则，通过规则间接实现治理目标，而敏捷治理强调去规则化，通过更灵活的政策直接以治理目标为导向。此范式下，各主体拥有更多裁量权，不再被动地遵守规则以避免法律责任，而是通过更加灵活的政策直接回应人类社会的根本需求^[21]。这要求政策制定者摆脱对僵化和过时规则的依赖，更直接地关注公众的即时需求和期望，深入理解技术如何服务于人类社会的根

本需求,同时也更加敏锐地捕捉到技术发展可能带来的问题和挑战。

2.2 去规则化发展:以核心价值和软法为引领

敏捷治理的去规则化指的是减少僵化规则在治理中的作用,转而采用更灵活、弹性的治理方式。这要求从宏观层面明确治理目标,将其分解为可实现的核心价值,以价值为导向进行治理。在微观层推动由硬法向软法的转变,以更有弹性和适应性的治理基础推动治理目标的实现。

规则导向指的是在治理中提出一系列明确的行为指引,行为人面对规则时只有“遵守或违反”两种选择。目标导向则是通过提出一系列价值,需要执法者在特定情境下根据一定的原则和背景综合判断^[22]。相较之下,目标导向提供了更大的灵活性和适应性,使法律体系能够更有效地适应AI背景下快速变化的社会和技术环境。目标导向的原则体系鼓励创新,避免了法律的僵化,促进公平和个性化的裁决。这种灵活的方式在面临新情况和无法预见的问题时,无须频繁修改法律文本,因此尤其适应于生成式AI的治理。

这一导向在2019年经济合作和发展组织(Organization for Economic Cooperation and Development)发布的人工智能治理原则的文件中有所体现。文件中提出一系列指导负责任的AI开发和使用的核心价值观和原则,旨在促进创新、信任和对人类福祉的尊重。这一系列原则包括包容性增长、可持续发展和福祉、人权和民主价值观、透明度和责任、技术鲁棒性和安全、公平和非歧视、利益相关者合作等。受该文件影响,以目标为导向设定AI治理核心价值也被各国广泛接纳。以英国2023年发布的《支持创新的人工智能监管方式》(A pro-innovation approach to AI regulation)政策文件为例,英国在AI的治理中,以促进发展和繁荣、增强公众对AI的信任、增强英国在全球AI领导地位为目标,设定了AI治理的5个价值导向:安全、安保与稳健性(safety, security and robustness),适当的透明度和可解释性(appropriate transparency and explainability),公平性(fairness),责任和治理(accountability and governance),争议性和救济(contestability and redress)。英国的治理框架以价值为基础构建,旨在确保AI的发展和应用既能带来社会经济收益,又不损害个人的权利和自由,同时提升公众对AI的信任度和接受度。英国政府在AI治理中秉

承敏捷原则和比例原则,确保干预措施与风险大小相匹配。因此,相较于遵循固定的法律法规,英国政府在治理时享有充足裁量权,同时秉持谦抑原则。

敏捷治理的软法化导向是相较于硬法而言的。硬法指的是正式法律法规,虽然在确立权利义务和行为及标准方面具有确定性和权威性,但其制定和修订过程相对缓慢,难以适应生成式AI快速迭代^[23]。相较于硬法,软法有以下4个优势:首先,不同于硬法需要遵守严格的立法流程,以政府指导建议、政策建议、行为准则、行业规范等为代表的软法制定具有灵活性和适应性;其次,不同于硬法具有普遍适用性,软法能够提供更具有针对性的解决方案,适应生成式AI的细分应用场景,并能随着技术进步和市场需求变化及时调整;再次,软法可以促进多元治理的实践,软法的制定通常需要政府与科研机构、行业协会、头部企业深度合作,更接近市场需求,更能促进产业发展并有效应对新兴风险;最后,软法具有宣示性和实验性,与硬法产生联动效果,一方面,宣示了政府立法态度和政策取向,为AI产业的适应提供时间,另一方面,为硬法的制定提供了试验场,政府和立法机关可以根据软法的实施效果和反馈,将验证过的、具有可持续性的风险管理和治理模式纳入法律体系。

软法在域外AI治理领域发挥重要的作用。以英国为例,其出台了包括指南、指导原则、标准、政府与私营部门协议、路线图等近70项软法作为AI治理依据,其中具有代表性并具有参考意义的文件如表1所示。

与依靠法律强制力和执法、司法等手段确保其实施的有效性的硬法不同,生成式AI的软法治理更加注重通过交流与合作建立社会各界的广泛共识,从而促进各方主动遵守。政府通过实施激励措施,如税收优惠、资金支持、政府采购优先等,鼓励企业遵守软法规定。与此同时,企业出于避免潜在的硬法规制带来更高的合规成本以及维护品牌声誉的考虑,也会主动调整经营管理策略以符合软法需求。从这个层面,在面临生成式AI快速迭代产生的未知风险和挑战时,软法可以提供有效的执行力。

3 敏捷治理与迭代流程

传统治理范式遵循线性流程(linear process),在政策的制定和执行时按照确定的顺序和阶段性路径执行,通常包括问题识别、政策设计、实施、监

表1 英国具有代表性的软法治理依据

软法名称	颁布机构及时间	类型	目的
《AI部门协议》	英国政府 2019.05	政府与私营部门协议	通过协议方式推动AI技术的商业化发展,同时确保技术创新的负责任和可持续
《AI路线图》	英国政府 2021.01	产业发展线路图	指导英国在AI领域的长期战略发展,强调研究、技术创新、人才培养和国际合作
《国家AI战略》	英国政府 2021.09	政府战略指南	通过该战略促进AI的商业应用,吸引国际投资,培养下一代科技人才
《AI与数据保护指引》	英国信息专员办公室 2023.03(更新)	政府机构颁布的使用指南	为各组织使用AI时处理数据的风险和挑战设定使用指南
《支持创新的AI监管方式》	英国政府 2023.03	白皮书	提出了英国AI监管的基本原则和治理框架
《AI标准》	英国标准协会 2024.01	协会标准	受政府和行业认可,制定涵盖AI设计、开发、应用等一系列标准,以确保技术的安全、可靠和伦理
《AI保障入门指南》	英国政府 2024.02	政府机构颁布的使用指南	支持组织更好地了解如何使用AI保障技术,以确保安全和负责任地开发和部署AI

督和评估等步骤。其僵化和缓慢的反应速度导致难以适应技术的快速迭代。相较之下,敏捷治理采用迭代的方法,允许政策制定和实施的过程中不断评估和调整,从而可以快速响应不断变化的需求。基于此,本节将分析迭代流程相较于线性流程的优势,并从技术支持和域外实践的角度探索迭代流程的建构模式。

3.1 从线性流程到迭代流程

传统治理范式依据固定的线性流程,这种方法旨在为政府行为带来明确性和可预测性,有助于确保政策的制定和执行的透明度。然而,此类范式在面对AI的快速迭代以及发展需求方面具有一定的局限性。

其一,明确的治理流程难以有效应对技术发展带来的时间敏感(time sensitivity)问题。在线性流程中,法律的制定有固定流程,时间相对较长。而AI技术,尤其是生成式AI,其发展和应用具有迅速迭代更新的特点,在短时间内就能实现显著的性能提升或功能扩展。如GPT-4在GPT-3.5发布后不到4个月时间即被推出,相比于GPT-3.5,其无论在模型的规模、生成质量还是理解复杂性方面都有显著的提升。因此,在政府通过线性流程制定并实施相关规范之前,生成式AI可能已经经历了更新迭代,拥有更加丰富的功能,并可能呈现出新的风险形态。其二,线性流程基于其结构化的特点导致创新空间受限。数据的爆发性增长对于AI的突破具有重要意义,传统AI往往依赖于较小的、特定的数据集进行训练,如图像识别系统等,而生成式AI需要大量数据来训练复杂的模型,其对数据

量以及数据多样性的需求远超传统AI。基于此前技术发展情况建立的数据安全、个人信息保护以及知识产权保护体系等可能妨碍数据汇聚融合,难以为AI发展提供丰富的数据基础,具有较高时间延展性的线性流程可能会成为我国AI发展的制约因素。

区别于传统治理的线性流程特征,迭代流程下的政策具有实验性、灵活性和适应性,通过不断计划、开发、测试、反馈、分析和改进,形成持续循环,每个阶段结束时需要立刻收集反馈并进行相应调整,以持续优化政策和实践。这种模式鼓励快速试错和即时改进,使政策的制定和执行过程能够快速适应新变化。敏捷治理通过短周期迭代,允许政策制定者和监管机构及时评估技术发展带来的新风险和挑战,同时也能快速整合新的科技进展和社会反馈,确保政策和监管措施始终反映当前的技术状态和社会价值观,促成生成式AI技术的快速健康发展。

3.2 去线性流程化：政策数字化与监管沙盒

迭代流程主要的实践形式是监管沙盒,其高效实施依赖于政策数字化(digital-ready policy)^[24]。政策数字化为精准决策、实时反馈和快速迭代提供了基础。通过数据分析和监控,政策制定者可以及时识别问题并调整策略,从而实现个性化和精准治理。通过引入数字化模型,监管沙盒允许在受控的环境中测试和优化政策,在快速变化的环境中通过迭代流程形成即时反应与应对策略,在促进发展的同时推动治理政策的适应性调整。

政策数字化在政府治理和市场经济的数字化转型背景下产生,旨在确保政策制定之初就全面考虑

数字化因素,为迭代流程的实现奠定技术基础。这一策略意在使政策适应数字时代的需要并预见未来发展,通过数字化手段整合政策设计、实施、评估等环节,提升政策的互操作性,确保系统、平台和组织间有效交流与合作。丹麦已将政策是否数字就绪作为评估立法的重要标准,新西兰也在试点将政策和法律编码化。政策数字化一方面可以与政策模拟模型结合,在政策实施之前进行政策模拟,评估政策的潜在效果和影响,例如,爱尔兰政府通过“智能经济创新政策模拟”让政府在政策实施前评估政策对爱尔兰经济的影响;另一方面,在政策实施后,政府可以借助实时数据分析、政策模拟等数字工具以及科研机构、行业协会、企业、信息平台等的实时反馈,促进政策的快速优化以及政策目标的高效实施,也为监管沙盒的应用提供了技术环境。

政策数字化的赋能下,监管沙盒被视为迭代流程的主要实践。迭代流程被用于应对快速发展和风险不确定的领域,一方面需要风险可控,另一方面需要政策高度的弹性和适应性。监管沙盒作为创新监管实践,其实验性和适应性与上述需求高度耦合。与传统模式依赖于事前设定规则和长期执行的流程不同,监管沙盒允许企业在限定范围的受控的环境中实验和测试新技术,同时政府可以实时监控、评估创新技术带来的风险和影响^[25]。这种方法的灵活性使AI技术的潜在问题可以在早期被识别和解决,解决了传统线性流程对快速技术变革反应滞后的问题。通过设定明确的实验期限和目标,监管沙盒保证了实验活动的目的性和可控性,为政策制定者提供了宝贵的实践经验和技术支持,帮助政府实验新型治理策略,形成更加精准有效的监管策略。

监管沙盒最初主要应用于金融科技行业,如新加坡(2016年)、英国(2016年)、加拿大(2017年)、阿联酋(2017年)的实践。但其实用性不限于此。实际上,监管沙盒的机制同样适用于生成AI领域,包括基础模型、应用模型以及更广泛的AI应用场景。这一点在欧盟委员会于2020年2月发布的《人工智能白皮书》中得到强调,该文件突出了将监管沙盒应用于AI治理的战略重要性。此外,欧盟《人工智能法》于2024年5月正式通过,其中设置“支持创新的措施”专章,以立法形式确立AI监管沙盒的应用。在此之前,欧盟轮值

主席国西班牙已批准建立第一个欧盟AI监管沙盒,使私营和公共部门组织能够在该法生效前根据其需求测试AI系统。以监管沙盒为代表的迭代流程治理将成为全球AI领域创新与合规并行发展的关键工具^[26]。

4 生成式AI敏捷治理的适用与构建

面对生成式AI引发的新兴的、不确定的治理难题,敏捷治理可以更好地满足其发展需要与治理需求。同时要注意到,敏捷治理与传统治理并非对立关系。在我国生成式AI治理模式的构建过程中,需要综合考虑风险未知性与风险程度、技术发展阶段等因素,在不断丰富敏捷治理工具箱的同时,充分发挥传统治理确定性和稳定性的优势,进一步促进生成式AI健康发展和规范应用。

4.1 敏捷治理在生成式AI治理中的适用

敏捷治理与传统治理的方法是一体两面、相辅相成的关系,共同构成一个综合治理体系。

敏捷治理虽然提供了必要的适应性和灵活性,但并非没有弱点,特别是在确定性和稳定性方面存在不足,难以完全满足生成式AI这一复杂领域的所有治理需求。如图2所示,相较于传统治理范式,敏捷治理更适用于确定性较低以及风险程度较低的领域。具体而言,在数据安全、个人隐私保护等确定性较高、立法和监管机制相对成熟的领域,传统治理相较于敏捷治理更具适用性。在这些领域引入敏捷治理,反而可能加剧市场的不确定性。同样,在涉及国家安全等重大风险的领域,任何治理上的疏漏都可能带来不可逆的后果,此时,传统治理由于其全面性和严格性而更具优势。相较而言,敏捷治理的适用性也较弱。因此,在生成式AI的治理实践中,需要综合考虑风险确定性、风险程度等多方面因素,有针对性地选择相应的治理手段。

敏捷治理适合应对生成式AI快速发展中的新挑战和未知风险,但对于已有明确法规和成熟治理机制的风险问题,则应优先考虑适用现有的传统治理体系^[27]。例如,针对生成式AI的数据安全问题,可以依托于《数据安全法》等法律规定中建立的数据安全风险评估、数据安全审查、数据安全应急处置机制等相对成熟的制度措施进行治理。但也需要承认,生成式AI可能带来的共同无知现象,导致立法者、政府和企业仍难以全面预测和评估可能产

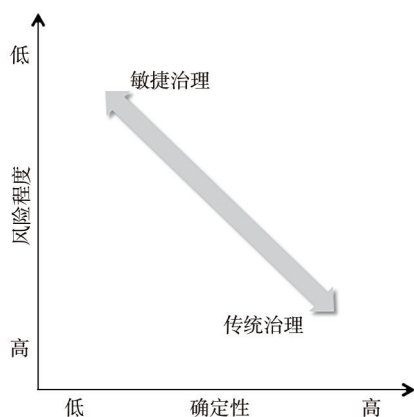


图2 生成式AI的治理范式选择

生的数据安全风险，在面对未知风险时，需要采取灵活适应的敏捷治理范式来识别和应对风险。在相对确定的领域应用敏捷治理会增强市场的不可测性，迫使企业放弃既有的合规体系来适应政府新的治理模式，进而投入更多的精力去适应不断变化的规则，而非专注于创新和发展^[28]。这种治理方式虽然意在提高适应性和响应速度，但在稳定性要求较高的领域会导致资源错配和效率下降。

敏捷治理适合应对生成式AI发展中风险程度较为可控的场景，但对于风险程度较高的领域，则仍需以严格管理要求打造高水平安全。在适用敏捷治理的过程中，不仅需要关注风险的确定程度，更需要做好高风险的防范控制。高风险领域中任何错误或疏漏都可能带来严重的后果。相较于传统治理范式强调事前的全面规划和严格的合规检查，敏捷治理虽然强调灵活性和快速适应变化的能力，但可能缺乏必要的安全保障和稳定性，因而其适用范围具有局限性。因此，为了建立适当的敏捷治理适用场景，需要从政府到企业建立一套全面的风险评估体系，根据风险程度的高低使用不同的治理模式。例如，《数据安全法》对数据实行分类分级保护，并对“重要数据”“核心数据”实施严格管理。类似地，欧盟《人工智能法》按照系统对用户和社会的潜在影响，将AI风险分为“不可接受风险”“高风险”“有限风险”和“最小风险”。特别是在高风险领域，企业在进入市场前需要履行包括透明度、可解释性、数据质量和人工监督等一系列严格义务^[29]，此类对高风险领域的严格管理显示出明显的传统治理范式的特征，也体现了传统治理与敏捷治理之间互为补充的关系^[30]。

4.2 中国生成式AI敏捷治理模式的构建

近年来，我国在积极探索构建敏捷治理模式，以适应科技创新发展的现实需要。在实践过程中，我国逐渐探索多主体参与的治理体系，积极统筹新兴领域发展与安全之间的关系^[31]。然而，针对生成AI技术的颠覆性突破，需要进一步深化多元主体间的合作共治，丰富AI治理“工具箱”，提高敏捷治理水平，进一步实现高质量发展和高水平安全的良性互动。

其一，在“多元参与”的基础上进一步探索“多元共治”。相较于多元参与，多元共治注重发挥各个主体间的主观能动性和各主体间的协同^[32]，强调公私合作、信息共享。伴随生成AI爆发式发展，我国多主体积极参与生成式AI的健康发展与风险应对。例如，政府部门坚持发展和安全并重，出台部门规章规范生成式AI的应用；行业组织发挥桥梁纽带作用，汇集企业有关规章规定的意见并反馈相关部门，发布有关标准指导生成式AI服务提供者提高安全管理水平；企业积极探索自治措施，通过企业AI治理平台推动建设开放、可信的AI创新生态。上述实践在一定程度上体现了多元主体参与治理的积极性，然而，各主体在参与AI治理过程中“各自为战”、缺乏互动，在前期治理工具的选择和后期治理经验的交流等方面均缺乏沟通渠道，未能形成敏捷互动，与形成敏捷治理“共治”模式还具有一定差距。未来，可以参考政策实验室，依托现有“国家新一代人工智能治理专业委员会”等组织，进一步成立人工智能治理工作组，广泛吸纳来自政府、行业协会、高校、科研院所、企业相关的专业人员，共同研究促进创新的制度措施和风险应对方案，增强多元主体之间的互动。该工作组应当在AI相关政策法规的制定中发挥协调作用，具体包括两方面：一方面，立足促进发展，定期召开研讨会，深入企业开展调研，研究适应科技发展的版权制度、数据产权制度等，探索促进AI发展的政策举措；另一方面，立足防范风险，鼓励企业投资AI风险检测工具和防御技术等，及时共享AI自主意识提高等造成极端风险的可能，并将具体情况向相关部门反馈，快速制定相关监管政策。同时，针对人工智能发展过程中亟待解决的具体问题，可以成立工作专班，组织具体领域政府部门、专家学者有针对性地探讨治理方案。

其二，在明确治理目标的基础上，由硬法与软

法协同发力。目前,我国已经就AI规范应用制定出台了一系列法律法规。一是构建了数据治理顶层架构,《网络安全法》《数据安全法》《个人信息保护法》等法律文件中明确了一系列有关数据治理的制度要求;二是明确了AI服务管理规定,出台算法推荐、深度合成、生成式AI服务管理规定,重点明确相关服务提供者的责任义务;三是规范AI伦理问题,公布了《新一代人工智能伦理规范》《关于加强科技伦理治理的意见》等系列文件,为从事AI相关活动的主体提供伦理指引。适应新时代的技术发展趋势,还需要探索构建法律法规、政策文件、行业规范等相互补充、相互配合的规则体系。在生成式AI快速发展初期,通过相关政策指导文件等具有指引性质的软法探索数据使用、知识产权归属等领域的创新规则,宣示政府促进发展的立法态度和政策取向,打通束缚AI发展的制度堵点;同时坚持问题导向原则,通过立法明确生成式AI在可能影响国家安全、可能影响人类安全及基本权利的领域的应用规范,形成基本的监督,防范可能产生的高风险问题。随着生成式AI在各个领域的深入应用,政府可以利用监管沙盒等敏捷治理手段在特定场景下探索并试验出有效的治理模式,并在这些治理模式被验证可靠后,通过行业标准、最佳实践、部门规章等将相应领域的治理经验程式化和固定化,为市场创造出稳定可预期的环境。在敏捷互动中,充分了解目前阶段AI的技术特点、潜在风险和社会影响,通过AI专门立法制定更精准和合适的法律制度,实现从敏捷治理到传统治理、从灵活到长期稳定治理的转变。

其三,在确立敏捷治理理念的基础上逐步丰富治理工具箱^[33]。敏捷治理的理念日渐融入我国有关新技术新应用新业态治理的政策或法律之中。如《新一代人工智能治理原则——发展负责任的人工智能》《关于加强科技伦理治理的意见》等文件均将敏捷治理作为治理原则之一;《生成式人工智能服务管理暂行办法》也规定“对生成式人工智能服务实行包容审慎和分类分级监管”。然而,目前我国关于敏捷治理的实践偏向于原则性规定,有待从多个方面丰富敏捷治理工具箱。具体到生成式AI治理方面,一是建立风险评估体系,提高治理模型选择的适应性。可以从政府到企业建立一套全面的风险评估体系,鼓励企业持续进行信息收集、分析和反馈,以精确评估生成式AI可能带来的风险,

确保政府采用适宜的治理模式应对技术进步,平衡创新推动和风险控制的需求。二是可以进一步提高数字政府水平^[34],建立健全大数据辅助科学决策机制,充分汇聚整合经济运行、区域概况、AI行业发展等多源数据资源,用信息化手段更好地辅助决策施政,在技术条件成熟时,还可以借鉴新西兰、丹麦、爱尔兰等国家的经验,试点利用数字化、智能化手段模拟政策法律实施后的效果,全面提升政府决策科学化水平。

5 结束语

本文通过分析生成式AI的发展挑战,探讨了敏捷治理模式的构建与应用。研究表明,敏捷治理模式具有多元共治、灵活治理依据和迭代流程的特点,能够有效满足技术创新与风险管理之间的平衡需求。国际案例分析进一步证实了敏捷治理在促进创新、提高政策适应性和响应速度方面的优势。中国在构建生成式AI敏捷治理模式时,应充分借鉴国际经验,结合国内技术发展和社会需求,灵活运用敏捷治理原则,以实现对新兴技术的高效治理。

敏捷治理范式并非要取代传统治理范式,而是作为一种补充,为治理体系提供更多的灵活性和创新空间。它不是一套固定不变的规则或模式,而是一种以原则为导向的治理思维方式,强调在不断变化的技术和社会环境中的适应性和迭代性。在具体应用时,应根据我国的实际技术发展水平、社会需求、法律法规和伦理标准等多方面因素,灵活选择和调整合适的治理模式和手段。通过这种动态平衡的方法,我们可以更好地应对生成式AI带来的风险,同时促进生成式AI在各个领域的健康发展和创新应用。

参考文献:

- [1] 郑南宁. 人工智能新时代[J]. 智能科学与技术学报, 2019, 1(1): 1-3. ZHENG N N. The new area of artificial intelligence[J]. Chinese Journal of Intelligence Science and Technology, 2019, 1(1): 1-3.
- [2] DRUCKER P F. Management challenges for the 21st century[M]. New York: HarperBusiness, 1999.
- [3] CALO R. Artificial intelligence policy: a premier and roadmap[J]. University of California, Davis Law Review, 2017, 51: 399-435.
- [4] 中国信息通信研究院, 中国科学院. 大模型治理蓝皮报告: 从规则走向实践[R]. 2023. CAICT, ICT. Large model governance blue paper report- from rules to practice[R]. 2023.

- [5] SCHWAB K, DAVIS N. Shaping the future of the fourth industrial revolution: a guide to building a better world[M]. Currency, 2018: 307-314.
- [6] World Economic Forum. Agile governance: reimagining policy-making in the fourth industrial revolution[A]. 2018.
- [7] KAAL W A, VERMEULEN E P M. How to regulate disruptive innovation - from facts to data[J]. SSRN Electronic Journal, 2017: 169-209.
- [8] 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[J]. 东方法学, 2023(4): 29-43.
- LIU Y H. On the three security risks and legal regulation of generative artificial intelligence[J]. Oriental Law, 2023(4): 29-43.
- [9] 张凌寒, 于琳. 从传统治理到敏捷治理: 生成式人工智能的治理范式革新[J]. 电子政务, 2023(9): 2-13.
- ZHANG L H, YU L. From traditional governance to agile governance: governance paradigm innovation of generative artificial intelligence[J]. E-Government, 2023(9): 2-13.
- [10] 薛澜, 赵静. 走向敏捷治理: 新兴产业发展与监管模式探究[J]. 中国行政管理, 2019(8): 28-34.
- XUE L, ZHAO J. Toward agile governance: the pattern of emerging industry development and regulation[J]. Chinese Public Administration, 2019(8): 28-34.
- [11] 丁晓东. 论算法的法律规制[J]. 中国社会科学, 2020(12): 138-159, 203.
- DING X D. On the legal regulation of algorithms[J]. Social Sciences in China, 2020(12): 138-159, 203.
- [12] 于文轩. ChatGPT 与敏捷治理[J]. 学海, 2023(2): 52-57.
- YU W X. ChatGPT and agile governance[J]. Academia Bimestris, 2023(2): 52-57.
- [13] FUI-HOON NAH F, ZHENG R L, CAI J Y, et al. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration[J]. Journal of Information Technology Case and Application Research, 2023, 25(3): 277-304.
- [14] EVAN J, OROY K. Prospects of AI: assessing potential, managing risks, and ethical frontiers[R]. EasyChair, 2024: 12681.
- [15] 陈天昊, 徐玮. 规范敏捷式监管: 金融数据开放的监管挑战与完善路径[J]. 治理研究, 2023, 39(6): 142-156, 160.
- CHEN T H, XU W. Normative agile regulation: regulatory challenges and improvement paths for financial data openness[J]. Governance Studies, 2023, 39(6): 142-156, 160.
- [16] VALLE-CRUZ D, CRIADO J I, SANDOVAL-ALMAZÁN R, et al. Assessing the public policy-cycle framework in the age of artificial intelligence: from agenda-setting to policy evaluation[J]. Government Information Quarterly, 2020, 37(4): 101509.
- [17] OLEJNICZAK K, BORKOWSKA-WASZAK S, DOMARADZKA-WIDŁA A, et al. Policy labs: the next frontier of policy design and evaluation[J]. Policy & Politics, 2020, 48(1): 89-110.
- [18] KIM S, WELLSTEAD A M, HEIKKILÄ T. Policy capacity and rise of data-based policy innovation labs[J]. Review of Policy Research, 2023, 40(3): 341-362.
- [19] KALENZI C. Artificial intelligence and blockchain: how should emerging technologies be governed?[J]. Frontiers in Research Metrics and Analytics, 2022, 7: 801549.
- [20] OZKAYA I. Application of large language models to software engineering tasks: opportunities, risks, and implications[J]. IEEE Software, 2023, 40(3): 4-8.
- [21] ANDERLJUNG M, BARNHART J, KORINEK A, et al. Frontier ai regulation: managing emerging risks to public safety[EB/OL]. arXiv preprint, 2023, arXiv:2307.03718.
- [22] 袁方成, 魏玉欣. 数字化转型何以助推敏捷治理? 以淮安“码”上议工程为分析对象[J]. 行政论坛, 2024, 30(1): 94-106.
- YUAN F C, WEI Y X. How digital transformation facilitates agile governance: take the discussion on “code” Project of Huai'an as the object of analysis[J]. Administrative Tribune, 2024, 30(1): 94-106.
- [23] WASISTIONO S, ANGGRIAINI W. Three paradigms in government[J]. International Journal of Kybmology, 2019, 4(2): 79-91.
- [24] MARCHANT G, GUIERREZ C I. Soft law 2.0: an agile and effective governance approach for artificial intelligence[J]. Minnesota Journal of Law, Science & Technology, 2022, 24(2): 375-424.
- [25] TRUBY J, BROWN R D, IBRAHIM I A, et al. A sandbox approach to regulating high-risk artificial intelligence applications[J]. European Journal of Risk Regulation, 2022, 13(2): 270-294.
- [26] ANTUNES H S, FREITAS P M, OLIVEIRA A L, et al. Multidisciplinary perspectives on artificial intelligence and the law[M]. Springer Nature, 2024: 441-457.
- [27] 张凌寒. 深度合成治理的逻辑更新与体系迭代: ChatGPT 等生成式人工智能治理的中国路径[J]. 法律科学(西北政法大学学报), 2023, 41(3): 38-51.
- ZHANG L H. Logic update and system iteration of deep synthetic governance: ChatGPT and other generative AI governance in China[J]. Science of Law (Journal of Northwest University of Political Science and Law), 2023, 41(3): 38-51.
- [28] 张欣. 生成式人工智能的算法治理挑战与治理型监管[J]. 现代法学, 2023, 45(3): 108-123.
- ZHANG X. Algorithmic governance challenges and governance supervision in generative artificial intelligence[J]. Modern Law Science, 2023, 45(3): 108-123.
- [29] 苏宇. 大型语言模型的法律风险与治理路径[J]. 法律科学(西北政法大学学报), 2024, 42(1): 76-88.
- SU Y. Legal risks and governance paths of large language models[J]. Science of Law (Journal of Northwest University of Political Science and Law), 2024, 42(1): 76-88.
- [30] HACKER P, ENGEL A, MAUER M. Regulating ChatGPT and other large generative AI models[C]//In Proceeding of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago US. New York: Association for Computing Machinery, 2023: 1112-1123.
- [31] 张凌寒. 中国需要一部怎样的《人工智能法》? 中国人工智能立法的基本逻辑与制度架构[J]. 法律科学(西北政法大学学报), 2024, 42(3): 3-17.
- ZHANG L H. What kind of “AI law” should China have? Basic logic and systemic structure of China's AI legislative framework[J]. Science of Law (Journal of Northwest University of Political Science and Law), 2024, 42(3): 3-17.
- [32] 李紫萱. 多元主体参与城市社区人工智能治理的协同路径: 基于敏捷治理理论[J]. 中国管理信息化, 2022, 25(18): 211-214.
- LI Z X. Collaborative path of multi-agents participating in artificial intelligence governance in urban communities—based on agile governance theory[J]. China Management Informationization, 2022, 25(18): 211-214.
- [33] 贾开, 赵静, 傅宏宇. 应对不确定性挑战: 算法敏捷治理的理论界定[J]. 图书情报知识, 2023, 40(1): 35-44.
- JIA K, ZHAO J, FU H Y. Dealing with challenges of uncertainty: theoretical definition of agile governance on algorithms[J]. Documentation, Information & Knowledge, 2023, 40(1): 35-44.
- [34] LI Y R, FAN Y Y, NIE L. Making governance agile: Exploring the role of artificial intelligence in China's local governance[J]. Public Policy and Administration, 2023: 09520767231188229.

[作者简介]



迟舜雨（1989- ），男，博士，中国政法大学法与经济学研究院助理教授，主要研究方向为资本市场、平台经济和科技治理。



杜安琪（1997- ），女，中国信息通信研究院助理工程师，主要研究方向为网络安全、数据安全、新技术新应用治理。