·科学论坛·

中国式治理现代化视域下的人工智能治理

张 辉[1,2] 曾 雄[2] 刘 鹏[3*]

1. 上海人工智能实验室，上海 200232
2. 清华大学 人工智能国际治理研究院，北京 100084
3. 河北省科学技术厅，石家庄 050021

[摘 要] 人工智能治理是国家治理现代化的重要组成部分。中国共产党第十八次全国代表大会以来，伴随着国家治理体系和治理能力现代化持续推进，以人工智能为代表的科技治理进入了一个新的历史发展阶段。本文运用“理念—体系—能力”综合性分析框架研究发现，中国特色的人工智能治理是一个以理念变革引领体系和能力变革的现代化过程。人工智能治理是政府、社会、市场等利益相关主体通过正式和非正式的制度安排，共同推动人工智能体系的创新、科研、生产及应用，并利用人工智能提升人类福利的综合性过程。在治理理念现代化的引导下，可持续推进治理体系现代化和治理能力现代化的历史进程。治理体系的构建和治理能力的建设，又为治理理念的创新与发展提供了基础性动力，最终形成中国特色的人工智能治理现代化图景。

[关键词] 人工智能治理；理念—体系—能力；共识性治理理念；适配性治理体系；效能性治理能力

人工智能治理已经成为全球范围内备受关注的议题。人工智能治理是国家治理现代化的重要组成部分。从国家治理现代化的内在要求来看，科技治理是国家治理核心组成部分，科技现代化与治理现代化相辅相成。其中，高新科技发展与治理，既是科技现代化的重要组成部分，也是治理现代化的重要基础。人工智能作为最具代表性的新兴科技和通用目的技术，理应被高度重视。做好人工智能治理，是运用人工智能赋能经济社会发展和数字政府治理的基础与前提。人工智能治理是指，政府、社会、市场等领域的利益相关主体通过正式和非正式的制度安排，共同推动人工智能体系的创新、科研、生产及应用，并利用人工智能提升人类福利；同时，识别、预防和应对人工智能技术创新和应用导致的政治经济社会风险与不良影响。在大力发展人工智能的同时，国际社会也开始强调人工智能伦理、负责任地发展人工智能以及安全利用人工智能，寻求社会发展与风险控制的平衡。世界诸国或地区，如欧盟、G20、经济合作与发展组织、美国、日本、俄罗斯等，也纷纷在 2019 年发布了各自的人工智能发展原则和治理准则，推动国际人工智能治理进入新的阶段。中国也于同年发布了《新一代人工智能治理原则——发展负责任的人工智能》。

然而，这些散落各处的治理原则与实践经验还无法跟上人工智能发展与扩散的步伐，人工智能治理问题日益突出。目前，国内外人工智能治理的研究处于刚刚起步阶段，对人工智能治理的分析大都

刘鹏 经济学博士，河北省科学技术厅综合规划处工作人员，主要从事科技创新战略规划、区域协同创新体系建设、高新技术转移转化等科技管理工作。

张辉 管理学博士，上海人工智能实验室青年研究员，清华大学人工智能国际治理研究院兼职研究员。研究领域包括人工智能治理、科技创新政策、技术经济、技术哲学。

收稿日期：2022-10-17；修回日期：2022-11-03

* 通信作者，Email：sysulp@126.com

从单一学科视角展开，来自哲学[1]、经济学[2]、社会学[3]、管理学[4]、法学[5]以及计算机[6]等学科的学者都从各自领域进行了一定的探讨；或者，从人工智能体系的单一模块领域探讨具体的技术治理议题，如数据治理[7]、算法治理[8]、伦理治理[9]、自动驾驶治理[10]、人脸识别治理[11]等；但对于什么是人工智能治理、为什么进行治理以及如何治理等还没有形成深刻的成果与共识，更没有整体性的认知和分析框架，制约了人工智能治理的发展[12,13]。因此，人工智能治理，亟需一个综合性分析框架，并加以指导其治理体系和治理能力的现代化实践。

1　人工智能治理的“理念—体系—能力”分析框架

人工智能治理是一项内嵌于国家治理的特殊活动，是国家治理在高新科技领域的具体治理实践。换言之，国家治理现代化的运作逻辑和构成要素，实质上对理解人工智能治理并指导人工智能治理实践起到了举足轻重的作用。“国家治理是国家政权的所有者、管理者和利益相关者等多元行动者在一个国家的范围内对社会公共事务的合作管理，其目的是增进公共利益维护公共秩序”[14]。而从系统科学的角度来看，国家治理现代化是以治理思想与治理理念为引领、以治理制度体系为基座、以治理能力建设与效能获取为支撑的完整性结构功能系统。其中，国家治理的三要素即为治理理念、治理体系和治理能力。

中国国家治理现代化理论强调，国家治理现代化理论中强调，理念、体系和能力是构成国家治理的三个基础性要素，也是分析人工智能治理的三个重要维度。在数据治理、算法治理、伦理治理等现有专项治理的理论解释的基础上，结合国家治理现代化理论，本文拟构建一个“理念—体系—能力”的三维度人工智能综合性治理的研究框架。从而，从人工智能的治理理念、治理体系和治理能力等三个层次或维度，对人工智能领域的国家治理现代化展开分析。在“理念—体系—能力”的三层次分析框架中，人工智能的治理现代化建设过程中，以人工智能治理理念为先导，引领治理体系的构建和治理能力的变革；进而，人工智能治理体系和治理能力具有反向支撑作用，治理体系的进一步完善和治理能力的再提升会后向式推动人工智能治理理念的更新和升华。与此同时，人工智能治理体系建构与治理能力提升是一个相辅相成的有机整体，前者是后者的结构性载体，后者是前者的现象级表象。

在国家治理现代化体系中，居于首当其冲位置的是治理理念，即回答“为什么治理”的问题，也即明确阐述人工智能治理活动必须考究的多方利益诉求和必须遵循的价值取向。因此，治理理念传承和沉淀在于国家文化与历史精神，是国民共同的目标追求，同时，根据不同的实践地域和治理场域，需要因地制宜、因时制宜的做出必要调试。起基础先导作用的治理理念，通过各种媒介会从认知、情感、意识或行动倾向等方面影响利益相关方，就决定了后者最终的行为选择和行动路径。治理体系重点回答的是“如何治理”的问题，具体而言，包括国家治理中各个相关领域中的体制机制、法律法规、政策标准等制度性基础，以及国家治理各类组织的功能定位、基本结构、运行规则、操作机制与策略[15,16]。治理能力在治理体系的基础上，回答“能否治理”的问题，即利用治理体系来管理相关实际事务的能力，主要由制度吸纳力、制度整合力、制度执行力等构成[17,18]，是促进治理体系发挥整体效能的能动因素。

结合国家治理现代化理论和人工智能治理实务来看，人工智能治理主要包括治理理念、治理体系和治理能力等三个基础性要素。其中，人工智能治理的目标追求和价值导向构成了人工智能治理的核心治理理念，包括人工智能治理在国家治理现代化中的总体定位和处理实际问题的治理原则。价值目标的设定通常会面临“鱼与熊掌不可兼得”的困境，进而需要治理主体进行“艰难”权衡与“艺术”决策。例如，发展效率和公平公正在后发国家的治理活动中往往难以兼顾，而决策者必须因地因时制宜地平衡二者的关系；而在具体治理活动中的目标选择上，平台治理者往往需要面对相对公平的统一定价和相对效率的多级定价之间选择。

现代化治理体系被认为是改革开放以来中国以经济建设为中心的第一次转型后的第二次伟大转型[19]。人工智能治理体系的构建则是在传统的科技治理体系的基础上，充分考虑人工智能的发展与规制的典型特征之后，构建的治理体系。人工智能的治理能力则是依托治理体系进行实际事务治理所展现出来的制度效能，因此包涵着多个不同的维度；具体而言，包括科技治理能力、组织动员能力、交流合作能力等各个方面。综合治理体系和治理能力的整体角度来看，治理能力既包括多治理主体的能力，如组织与群体能力、个体性能力；也包括多层次的治理能力，如全球治理能力、国家治理能力、地区治理

能力、基层治理能力和社会公众与一线人员的治理能力等等；更包括治理过程中展现的能力，如源创新能力、供应链安全能力、产业链完整性能力、引进消化吸收再创新能力等等。

因此，本文提出人工智能治理的“理念—体系—能力”三位一体综合性分析框架（图 1）。其中，以理念为先导，以体系为载体，以能力为保障。正如政府战略管理的“三圈理论”指出的，相对于能力、支持而言，价值最为根本，是第一位的，是决策优先考虑的因素，是决策的前提条件[20]。人工智能治理的现代化理念在继承新兴科技治理理念的基础上，尊重人工智能的技术发展规律，引领人工智能的治理体系构建与完善和治理能力的获取与提升。人工智能治理体系通过具象化治理理念，并嵌入到具体技术实践场景中，获取并提升治理能力，进而丰富并升华治理理念。三位一体的基础性治理框架构成了新时代中国特色的人工智能治理谱系，也决定了中国人工智能治理特色发展的方向路径。

2 人工智能治理的治理理念现代化

启动人工智能治理，并在具体场景中加以落实，已经逐渐成为各界的共识。与此同时，各类主体针对各自的场景需求，给出了多重治理理念。在共识性治理理念方面，我国于 2019 年发布了《新一代人工智能治理原则——发展负责任的人工智能》，本质上为我们提供了共识性治理理念，并将之体系化，形成以“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、敏捷治理”等八条基本治理原则。该原则体系与其他代表性国家或地区的治理理念相比，更具包容平衡性。

从人工智能治理的某个维度或应用领域来分析治理的现状与问题，只有少部分学者提出人工智能

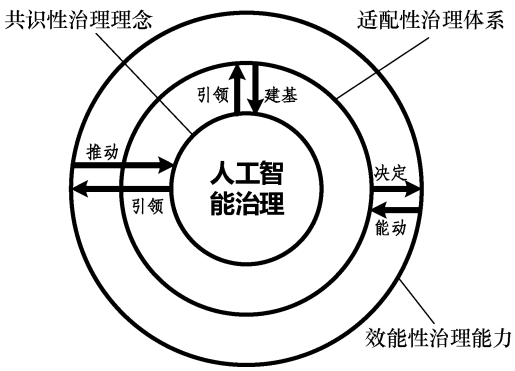

图 1 人工智能治理的“理念—体系—能力”综合性分析框架

综合治理的治理理念，其构想仍然缺乏适用性[21]。在实践中，新加坡已经率先展开尝试，其提出的人工治理框架被世界经济论坛认为是目前人工智能治理的最佳实践之一。该框架凝练出了两个治理原则，包括：（1）人工智能决策必须是可解释的、透明的且公平的；（2）人工智能系统必须是以人为中心的。一方面要对人工智能的发展保持支持和鼓励的基本态度；另一方面也要对人工智能发展的不确定性、潜在风险和负面影响给予充分的关注，在确保人工智能有序发展和安全可控的同时，着力防止人工智能的滥用。因此，对人工智能的发展我们应该保持包容审慎的基本态度，在确保人工智能安全和平等的底线基础上，祛除限制人工智能产业发展的制度束缚，利用人工智能赋能经济、社会和生态环境的可持续发展。换言之，人工智能治理理念反应了我国国家内部和社会公众对“人工智能发展与治理或中国科技治理”形成的一致而广泛的共识。

2.1 保证人工智能技术安全

保证技术安全是人工智能治理的底线目标，也是坚持“人民至上”和“科技以人为本”的基本出发点。习近平总书记在 2021 年 11 月的主持召开的中央政治局会议强调，加快提升生物安全、网络安全、数据安全、人工智能安全等领域的治理能力。且不谈人工智能能给经济社会发展带来多大的价值，使用人工智能应该首要保证不会对人类生命、健康以及财产等带来严重损害。就如新药物的开发与投入使用，药物的效果要经过层层的临床验证，确保没有健康问题，才能批准上市。在没有技术安全缺乏保证的情况下，不能仓促应用人工智能。或者虽有允许范围内的副作用或者潜在安全风险，明确告知其使用的用途、范围和方式等，避免误用。

2.2 维护公民尊严与平等

在保证人类身体和财产安全的基础上，人工智能治理需要维护公民精神上的尊严与平等，防止人工智能技术带来性别、种族、职业、年龄、地域、学历和兴趣爱好等方面的歧视，带来精神伤害和权利损失，促进社会公平。党的十八大以来，党和国家领导人围绕人权发表重要论述，结合中国历史和现实深刻阐明：“尊重和保障人权是中国共产党人的不懈追求。”

2.3 人工智能赋能经济发展

习近平总书记指出“人工智能是新一轮科技革命和产业变革的重要驱动力量”，强调“我们要深入把握新一代人工智能发展的特点，加强人工智能和

产业发展融合，为高质量发展提供新动能”。中华人民共和国科学技术部（以下简称“科技部”）、教育部、工业和信息化部等 6 部门联合发布《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》，统筹人工智能场景创新；科技部发布《关于支持建设新一代人工智能示范应用场景的通知》，支持建设包括智慧农场、智能港口在内的 10 个人工智能示范应用场景……近期，助力培育人工智能应用场景的政策措施接连出台，为牵引推动人工智能落地营造了良好的政策环境。人工智能跟电力一样，是一种通用型技术，在大规模生产的同时，能够根据用户需求提供个性化的产品与服务，便利人们的生活。在保证人工智能技术安全与公平的基础上，人工智能治理的一个现实目标是减少阻碍人工智能技术发展的不利因素，推动技术的广泛利用，使得更多生产部门和人口可以享受技术带来的红利。

2.4　人工智能促进可持续发展

站在人类社会的更高层面上，人工智能技术对于应对人类的重大挑战如气候变化、环境污染、传染病扩散等方面具有重大的潜力。这个层次的目标更多针对公共利益，因而需要公共部门提供更多的激励才能使得更多主体投入相关技术的研发与应用。同时，许多企业为了营造负责任创新的形象以及开辟新的竞争赛道，也可能积极利用人工智能来应对能源、食品、水资源等方面的挑战。“人类命运共同体”意识已经得到了中国人民的普遍认同，也逐渐在世界范围内得到承认，旨在追求本国利益时兼顾他国合理关切，在谋求本国发展中促进各国共同发展，人工智能发展和治理理应促进人类社会的可持续发展。

3　人工智能治理的治理体系现代化

人工智能治理体系的现代化建设在于具象化其治理理念，使之形成具体的治理结构和治理组织主体，进而夯实治理能力获取与提升的结构性基础。

3.1　人工智能治理的宏观制度体系化

国家规划是引导人工智能发展方向的重要工具，是一个国家体现人工智能治理价值的重要载体，不仅要规划技术和产业发展的目标和进程，也需对如何负责任开发与应用提出要求。例如，中国在 2017 年发布的《新一代人工智能发展规划》不仅明确了产业发展的路线图，也对加强人工智能治理研究与实践提出了重点要求。法律是最具权威的治理工具。法律制定人工智能技术开发、应用和开发的基本规则，明确价值链条上各主体的权利与义务，对治理主体形成强制压力。市场竞争是无形之手。充分市场的调节作用，将使企业和产品优胜劣汰，但市场也有可能逆向选择，需要国家规划的宏观指引。

伦理规范虽然缺乏强制力，但却是最容易使用并引导技术设计方向的治理工具。2016 年以来全球 80 多个人工智能伦理政策文件，发现不仅公共部门强调利用伦理政策文件来进行干预，私有企业也积极发布伦理原则来体现社会责任和领导力[22]。企业在市场竞争和社会责任双重压力下形成的内部治理规范，是人工智能治理的重要一环。新加坡的人工治理框架提出了组织在人工智能治理中四个具体的考虑维度：组织内部治理结构和方法、人工智能决策系统中人类的介入程度、执行管理以及参与者的互动和沟通，具有很大的启示。科技伦理委员会是人工智能企业履行科技伦理责任的基础性治理体系[22]。人工智能企业应该成立伦理委员会，制定人工智能伦理相关的内部标准与流程，并基于此对人工智能相关业务、产品或服务进行伦理审查，以识别、预防、消除人工智能相关产品或服务在安全、公平、隐私等方面的伦理风险。除此之外，对于人工智能开发中的数据、算法、系统、部署等多个环节的重点人员，企业应加强对其合规、伦理素养的培训。

3.2　人工智能治理的治理主体多元化

人工智能研发、应用与扩散中涉及到多个异质主体的权利与责任。因此，围绕着人工智能治理议题，本文梳理了其核心治理主体和外围治理主体，并明确各个治理主体的定位与治理职责。多元治理主体在人工智能社会技术系统中拥有不同的技术能力和产业链生态位，进而拥有不同的权威、资源、利益与限制；而通过各种正式与非正式渠道不断博弈平衡，多元主体可以构成人工智能治理的治理复合体。

“治理”强调多元主体对共同事务的管理，其中的权力运行不是自上而下的单向过程，而是上下互动或者水平沟通的双向过程。当前，人工智能正推动着不同治理主体角色的转变。例如，数据隐私保护条例的出台涉及数据生成者（用户）、数据聚合者（使用人工智能的平台企业）、数据使用者（研发机构）和数据监管者（政府及其他）等多方利益主体之

间的博弈和互动，各方都应当具有人工智能治理的知识合法性或参与合法性。上述技术发展路径和商业模式，同样决定了人工智能治理与传统技术治理框架存在诸多不同。在传统的治理框架中，政府通常是治理的核心，保有对社会（即各类非政府主体）的引导控制能力。因此，人工智能治理应该构建由人工智能企业（技术提供者或技术使用者）、公众（技术使用者）、高校、科研机构、政府部门、社会团体等共同组成的治理主体集合，明确权责的归属，有效地实现不同治理主体之间的灵活互动和敏捷沟通，从而更加高效地应对人工智能带来的多重治理挑战。

3.3 人工智能治理的治理对象明细化

人工智能包含众多要素、技术和场景，若不加区别地把整个人工智能作为治理对象，将会造成治理问题的“失焦”。在人工智能技术的研发、生产、应用和产生影响的过程中，既涉及人工智能算法和应用的特殊性问题，也包括普遍性的基础问题，如数据治理和个人信息保护，它们不仅是人工智能发展中面临的问题，也是许多其他数字技术和商业模式发展所需解决的问题。综合来看，人工智能治理的治理对象按治理场域可分为数据、信息、算法、算力、场景和技术外部环境。

数据是真实世界的记录，而信息是对社会主体（个体或组织）有意义或价值的数据。因此，并不是所有的数据都能产生有效信息，需要对数据和信息有一定的区别和对待，才能更加精准治理。数据治理必须兼顾数据保护和数据利用两个方面。在信息层面，很大部分数据的使用涉及到个人、企业和政府等组织的敏感信息，这些信息既是一种资产也是一种权利，个体和组织有权利保护自己的信息不被外人轻易获取。在《数据安全法》的基础上，《个人信息保护法》等法律法规可对信息层面提供保护和利用规则。

新一代人工智能算法治理的实质是在算法稳定性和算法安全性上取得的平衡。其中，人工智能算法的稳定性是指算法不会随着其他因素而改变其运行过程，算法的性能指标能够保持在一个合理变化的范围内。对于非数据驱动型算法而言，算法复杂度往往是提前可以预知的，其稳定性一般较好；而对于数据驱动型算法而言，算法的稳定性则取决于训练数据、测试数据和真实数据之间的相似度。人工智能算法的安全性是指算法处于不易受到攻击、算法模型参数的传输安全与不易泄露、算法运行过程安全可控等等状态。由于人工智能的安全治理属于一个全面系统性工程，算法安全性和数据安全性往往绑定在一起。因此，针对不同的算法特点，需要不同的规则和技术，在基础设施、算法、和数据安全等领域加大研发投入，发展出更加透明、可解释、安全和稳定的人工智能系统。

人工智能的算力系统是人工智能技术体系的根基所在，算力治理在于在核心基础软硬件基础设施的自主可控和技术创新需求的高速发展之间保持平衡。在承接上一轮中长期科技发展规划的基础上，以人工智能的大规模应用为契机，发挥经济外部需求对技术创新的拉动作用，倒逼底层技术的软硬件基础设施建设与模块创新，打破发达国家的技术垄断，最终形成整体性的系统创新升级。同时，为了应对旺盛的人工智能应用需求，科学合理地布局算力中心，并构建算力系统的动态调度网络。在保证人工智能产业以及相关应用产业高速发展的前提下，对算力系统的核心基础软硬件基础设施及其技术构件，通过激励自主创新和加强基础研究投入，实现有规划分批次的自有可控的核心基础软硬件的国有替代。最终，保障人工智能产业的健康可持续发展和科技安全。

人工智能是通用型技术，可大规模地应用到多样性的场景，需要结合不同场景的特征、需求和规则，形成“场景驱动”的治理体系。技术所引发的风险等级由风险发生的概率与风险一旦发生所引发的后果共同塑造，这与不同的应用场景高度相关。我们可以按照风险发生的概率和风险一旦发生其后果的严重程度将人工智能技术所引发的风险等级划分为四个等级：(1) I 级为发生的概率高且一旦发生其后果尤为严重的风险，此类风险需要优先治理，例如辅助医疗；(2) II 级为发生的概率高但其后果不算严重的风险，此类风险要及时治理，例如平台未征得用户同意进行数据收集造成的对个人隐私的侵犯，如人脸识别；(3) III 级为发生的概率低但一旦发生其后果尤为严重的风险，此类风险需要预见治理，例如自动驾驶；(4) IV 级为发生的概率低且即使发生其后果也不严重的风险，此类风险可以通过激发主体的志愿意识进行治理，如智能家居。

4 人工智能治理的治理能力现代化

人工智能治理能力现代化过程实则伴生于治理

理念达成的现代化和治理体系构建的现代化历程。面对国际国内的人工智能治理需求，快速达成价值共识，形成治理目标；进而构建治理体系并形成治理共同体，在具体的技术应用场域中获取并提升人工智能治理能力，是人工智能治理能力的典型现代化过程。具体而言，现代化的人工智能治理能力包括价值共识的快速形成能力、治理主体的合理分工与统筹兼顾能力、迭代优化治理能力和治理工具的科学合理使用的能力。

4.1　价值共识的快速形成能力

治理主体的多元化不仅催生了形态各异的治理体系，也使得人工智能治理的价值主张集合纷繁复杂。能够快速凝结价值共识，是人工智能治理能力现代化的重要体现。价值共治的形成，在于多元治理主体在秉承人类共同的或者朴素性的公序良俗理念，在充分沟通和互动的基础上，求取彼此价值主张集合的最大公约数。而当出现价值主张的冲突局面时，能够换位思考，在共治理念的驱动下，彼此做出必要的妥协而争取多赢局面。

价值共识的快速形成能力存在于两个范畴。在国际范围内，作为现代化治理的最大主体的国家实体，必须能够及时回应域外治理理念中的潜在冲突。世界诸国的发展阶段不同，发展历史和文化底蕴也不仅相同，使得彼此之间的人工智能治理理念各异。国家主体需要在国际层面遵循和平共处五项基本原则，形成适合自身发展的人工智能治理理念。在国内范围内，需要构建开明的发声媒介和畅通的沟通渠道，引导多元治理主体的治理理念的阐述与互动，进而快速迭代，推行敏捷治理原则在价值共识的形成过程中实际落地。

4.2　合理分工与统筹兼顾能力

多层次的治理体系、多元化的治理主体、明细化的治理对象都需要治理主体之间合理分配治理技能。梳理人工智能治理主体的价值分工，明确各类主体的优势与劣势，取长补短，形成优势互补，是人工智能治理的内在要求。合理的价值分工能够落地人工智能治理的治理理念，同时释放治理体系的效率和潜能。充分利用不同治理主体的治理能力，兼顾治理成本和治理效率，运用各类人工智能产品或服务赋能公共治理组织、治理群体与个体，同时，引导政府部门运用智能工具赋权其他相关主体，生成协同共治的治理态势，并形成治理合力，使得社会总福利最大化。

4.3　迭代优化治理能力

人工智能治理主体的理性偏好和治理诉求会随着技术发展和社会进步而保持变化，在既定的治理理念和治理体系的基础上，实施跟进经济社会的变化，厘清人工智能治理的实际需求，形成一种迭代优化的治理思维和治理能力。新一代人工智能的技术优势在于大数据驱动、云计算能力的大幅度提升等技术要素的积累，而这种积累达到一定的阈值之后会出现技术涌现现象，催生出新型的技术产品或服务，如赛博空间、智慧城市、元宇宙等新型技术场景，继而使得人工智能技术的社会嵌入程度越发宽泛而深入。与之相伴的技术风险也会随着这一进程而不断增大，这就需要实时调查并展开分析，进而保持治理理念的与时俱进和治理体系的迭代优化，形成因时制宜和因地制宜的效能型治理能力，释放中国特色现代化的治理潜能和中国民主集中决策的敏捷治理优势。

4.4　治理工具的科学运用能力

治理工具是治理主体用来解决治理问题的途径、方法和手段。不同的工具有不同的优势和局限，在不同尺度、场景中发挥不同的功效。在宏观尺度上，法律是治理最有强制力的工具，但法律制定一般无法跟上人工智能迅速变化的节奏，普适性和原则性较强的法律条款也难以满足人工智能许多个性化应用场景的需求。治理宣言、技术标准、行为规范、国际倡议等也逐渐被纳入人工智能治理工具的范畴之中，并根据具体的治理问题和治理需求加以利用，实现治理工具的多样化。

在中观尺度上，社会实验是一种有效的尝试。人工智能社会实验可以选取城市、农村、企业、医院、学校、政府机构等不同领域的真实应用场景，设立实验组和对照组进行长时间周期、跨空间区域、多学科综合的介入式观测，围绕风险认知、利益获得、价值形塑、组织变革、制度变迁、政策回应等测量指标，进行科学测量，对人工智能的治理工具以及社会影响进行综合性科学循证研究，为更大范围的人工智能治理提供科学的、第一手的理论参考、实践经验和技术规范。

在微观尺度，拥有更多的工具组合，主要包括技术标准，组织内部治理规范和监管科技。技术标准包括基础技术标准、产品标准、工艺标准、检测试验方法标准，及安全、卫生、环保标准等。标准可以通过规范产品的规格、可解释性、鲁棒性和故障安全设

计等特征影响特定人工智能系统的开发和部署。标准还可以通过规范开发流程影响人工智能研究、开发和部署的大环境。标准的建立、传播和执行可以在研究人员、研发机构和政府之间建立信任，并可以在全球范围内起到传播最佳实践的作用。随着越来越多的行业代表开始逐渐加入到标准制定的过程之中，未来标准（包括地方性、区域性、国际性等不同层级）将能够更好地对人工智能产业发展形成良性约束。

5 总 结

人工智能治理的现代化治理框架应当是包括治理理念的创新与发展、治理体系的建构与巩固、治理能力（包括治理工具、治理手段等）的获取与提升等三维现代化维度。治理现代化框架为人工智能治理提供了“理性坐标系”，通过精准定位治理主体和理性维度，能够为“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、敏捷治理”等治理原则和理念提供足够的执行空间，明确多元治理主体之间的分工与合作机制，确立人工智能治理的治理对象、治理方向和治理路径，最终形成人工智能治理的“理念—体系—能力”的现代化图景。

本文在八项基本治理原则的基础上，细化了人工智能治理理念，即保证人工智能技术安全、维护公民尊严与平等、人工智能赋能经济发展、人工智能促进可持续发展等四项更加具体可行的方案。而随着国家发展和人民物质精神文明水平的提升，我们更需要在上述治理分析框架的指导下更新适应发展的治理理念。进而从宏观制度体系化、中观层面的治理主体多元化和微观层面的治理对象明细化角度阐述了人工智能治理的现代化体系构建。最后围绕人工智能治理的治理能力现代化建设，着重探讨了价值共识的快速形成能力、迭代优化的治理能力和科学运用适宜的治理工具的能力建设。未来研究将围绕这人工智能治理理念更新、人工智能治理的体系现代化过程和结构特征、人工智能治理的现代化能力获取和效能生产展开分析。

参 考 文 献

[1] 程海东，王以梁，侯沐辰. 人工智能的不确定性及其治理探究. 自然辩证法研究，2020，36(2)：36—41.

[2] 李超，艾慧. 人工智能对我国就业创造效应和破坏效应研究——基于政治经济学角度研究. 当代经济，2022，39(9)：3—8.

[3] 管其平. 智能社会学：智能时代社会学研究的新方向. 华南理工大学学报(社会科学版)，2022，24(3)：1—9.

[4] 陈中飞，汪锋，董明放，等. 管理科学部经济科学学科人工智能指派与分类评审效果分析. 中国科学基金，2022，36(5)：819—824.

[5] 何邦武. 数字法学视野下的网络空间治理. 中国法学，2022(4)：74—91.

[6] 徐宗本. 人工智能的10个重大数理基础问题. 中国科学：信息科学，2021，51(12)：1967—1978.

[7] 孟小峰，刘立新. 区块链与数据治理. 中国科学基金，2020，34(1)：12—17.

[8] 谢康，夏正豪，肖静华. 大数据成为现实生产要素的企业实现机制：产品创新视角. 中国工业经济，2020(5)：42—60.

[9] 陈小平. 人工智能中的封闭性和强封闭性——现有成果的能力边界、应用条件和伦理风险. 智能系统学报，2020，15(1)：114—120.

[10] 张辉，梁正. 自动驾驶“单车智能”模式的发展困境与应对. 齐鲁学刊，2021(6)：81—89.

[11] 曾雄，梁正，张辉. 人脸识别治理的国际经验与中国策略. 电子政务，2021(9)：105—116.

[12] 陈杰，樊邦奎，邓方，等. 智能群系统的衍化与协同——第252期双清论坛学术综述. 中国科学基金，2021，35(4)：604—610.

[13] Leal Filho W，Azul A，Brandli L，et al. Partnerships for the Goals// Leal Filho W，eds. Encyclopedia of the UN Sustainable Development Goals. Springer Cham，2020. https://doi.org/10.1007/978-3-319-95963-4.

[14] 何增科. 理解国家治理及其现代化. 马克思主义与现实，2014(1)：11—15.

[15] 薛澜. 顶层设计与泥泞前行：中国国家治理现代化之路. 公共管理学报，2014，11(4)：1—6，139.

[16] 张树华，王阳亮. 制度、体制与机制：对国家治理体系的系统分析. 管理世界，2022，38(1)：107—118.

[17] 宋世明. 坚持在法治轨道上推进国家治理体系和治理能力现代化. 中国政法大学学报，2021(3)：19—31.

[18] 杨光斌. 关于国家治理能力的一般理论——探索世界政治(比较政治)研究的新范式. 教学与研究，2017(1)：5—22.

[19] 胡鞍钢，王绍光，周建明. 第二次转型 国家制度建设. 2版. 北京：清华大学出版社，2009.

[20] Moore MH. Creating public value：strategic management in government. Cambridge，Mass.：Harvard University Press，1995.

[21] Stanford University HAI. 2022 AI Index Report. https://hai.stanford.edu/ai-index-2022.

[22] 曹建峰. 人工智能治理：从科技中心主义到科技人文协作. 上海师范大学学报(哲学社会科学版)，2020，49(5)：98—107.

Artificial Intelligence Governance From the Perspective of Modernization of Chinese Style Governance

Hui Zhang[1,2]　Xiong Zeng[2]　Peng Liu[3*]

1. *Shanghai Artificial Intelligence Laboratory*, *Shanghai* 200232
2. *Institute for AI International Governance of Tsinghua University*, *Beijing* 100084
3. *Hebei Provincial Department of Science and Technology*, *Shijiazhuang* 050021

Abstract　Artificial intelligence governance is an important part of the national governance modernization. Since the 18th National Congress of the Communist Party of China, science and technology governance represented by artificial intelligence has entered a new stage of historical development with the continuous advancement of the modernization of the national governance system and governance capabilities. Based on the comprehensive analysis framework of "concept-system-capability", this paper finds that artificial intelligence governance with Chinese characteristics is a modernization process of system and capability transformation led by concept transformation. Artificial intelligence governance is a comprehensive process in which the government, society, market and other stakeholders jointly promote the innovation, research, production and application of artificial intelligence system through formal and informal institutional arrangements, and use artificial intelligence to improve human welfare. Under the guidance of the modernization of governance concept, the historical process of the modernization of governance system and capability should be continuously promoted. And the construction of the governance system and the governance capacity provide the basic power for the innovation and development of the governance concept. Finally, a modern vision of artificial intelligence governance with Chinese characteristics will be formed.

Keywords　artificial intelligence governance; concept-system-capability; consensual governance philosophy; adaptive governance system; effective and efficient governance ability

（责任编辑　崔国增　姜钧译）

* Corresponding Author, Email: sysulp@126.com